# UNCERTAINTY AND DECISION-MAKING IN THE HUMAN BRAIN

Iris M. D. Vilares

Dissertation presented to obtain the

Ph.D degree in Biology | Neuroscience

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Research work coordinated by:

FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

Champalimaud Foundation

Oeiras,

December, 2013

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

*To my family*

**APOIO FINANCEIRO/FINANCIAL SUPPORT**

Apoio financeiro da FCT e do FSE no âmbito do Quadro Comunitário de Apoio, bolsa de doutoramento n° SFRH/BD/33272/2007. O ultimo ano foi financiado pelo National Institute of Health, bolsa n° 2P01NS044393.

Financial support for this thesis was provided by FCT and FSE through the Quadro Comunitário de Apoio, doctoral fellowship n° SFRH/BD/33272/2007. The last year was provided by the National Institute of Health, grant n° 2P01NS044393.

**ACKNOWLEDGEMENTS**

Many people have helped me through this path, and I am immensely grateful to all of them. I would like to thank:

My supervisor, Dr. Konrad Kording. Thank you for being an amazing supervisor, for always being there to help me. You developed in me a love for science and always believed in my scientific skills, sometimes even more than myself. Thank you for your guidance and support at all times. For always feeling welcomed to go to you for any question, be it scientific or not. Thank you for an incredible mentorship. You are my scientific role model.

My internal supervisor, Dr. Rui Costa. Thank you for your enthusiasm and support at all times. Thank you for all the help you gave me in taking care of bureaucracy, especially this past month, when almost every day I was sending you an email. Your constant smile and optimism are a source of positive energy.

The International Neuroscience Doctoral Program, or INDP (previous Programa Gulbenkian Champalimaud em Neurociências) for giving me the possibility of doing my PhD, and allowing me to develop my research at Northwestern University. I and my colleagues were the “first batch” of this amazing program, and it has been a very interesting and rewarding experience. We had one year of full-time classes at the Instituto Gulbenkian de Ciencia (IGC), in which we could interact and learn with successful scientists from all over the world; and then we had the luxury to have 4 years of funding to go and do our research anywhere. Thank you to Dr. Zachary Mainen for embarking with us in this journey and being the first director of INDP. Thank you as well for your support and love for science. Thank you to the Fundação Champalimaud, the IGC and the Fundação para

a Ciência e Tecnologia (FCT), for funding INDP and giving institutional and monetary support. A special thanks to IGC for kindly being our host institution when the Champalimaud Foundation was still being built; to Fundação Champalimaud for institutional support and organizing retreats througout the years where I could discuss and present my work; and to FCT for providing me financial support (grant n° SFRH/BD/33272/2007).

My thesis committee members, Dr. Jorge Carneiro and Dr. Rosalina Fonseca. Thank you for your questions and for enabling me to look at my results through different perspectives. Thank you as well for your support, friendliness and friendship throughout the years.

The doctors at Northwestern Memorial hospital and the Rehabilitation Institute of Chicago, for showing me how rewarding and stimulating it can be to work with patients, for helping me recruit patients and for promoting the combination of science and motor rehabilitation. I would specifically like to thank Dr. Tania Simuni, Dr. Santiago Toledo and Dr. Deborah Gaebler.

The people that were always there to help me with paperwork, reimbursement, visa issues or whatever else came along. Namely, I would like to thank: Manuela Cordeiro, Ana Portocarrero, Teresa Carona, Alexandra Piedade, Lance Adams, Robin Millard, Paul Hixon, Kathryin Niedner, and, of course, Debbie Znajda Kitterman.

Hugo Fernandes, both colleague and friend. We started our PhD together and, if everything goes well, we will also finish it at the same time. Thank you for being there through the good and bad times. Thank you for basically helping me with all my projects, either through discussions, suggestions or even help making the figures. Thank you for giving me the

extra bit of motivation that helped me to go across the ocean for my PhD. I don't think this thesis would have existed without you.

Mark Albert, ex-colleague and still good friend. Thank you for making a great lab environment. Thank you for the discussions we had, the help you gave me with my projects, and the useful feedback you gave to my thesis. Thank you for always being there for me, both to celebrate with me when things were going well, and as a shoulder to rely on when things were difficult, both inside and outside the lab. Thank you for your positivism, energy, and for being such a great and trustworthy friend.

The other past and present members of the Kording lab, including Ben Walker, Daniel Acuña, Max Berniker, Steve Antos, Pat Lawlor, Ted Cybulski, Josh Glaser, Rebecca Moore, Paul Wanda and Kunlin Wei. A special thanks to Greg Dam, who first taught me how to use Matlab; to Ian Stevenson, whose capacity to write code at a super-natural speed still amazes me; and to Pavan Ramkumar: thank you for your support in writing the thesis (including making me company in coffee shops!), your friendliness and enthusiasm, and your friendship, that has been a source of positive energy in these last months. In general, thank you to all the members of the Kording lab for making the lab a friendly and scientifically stimulating place to work.

My colleagues from the INDP program, including Patricio Simoes, Patricia Correia, Rodrigo Abreu, Maria InêsVicente, Margarida Agrochão, Pedro Ferreira and Isabel Henriques. A special thanks to José Fernandes, who has helped me a lot in these last months, advising me on labs, thesis writing and paperwork; and to Mariana Cardoso, for being such a good friend and always always being there.

The whole range of amazing people that I have met in Chicago. Specifically, I would like to thank: Inês Domingues, who gave me food, shelter and friends when I arrived to Chicago; Lisa Morselli, Fanny Delebecque; Sandra Maximiano and Brenda Lopez Silva, who gave me their friendship from the very beginning and still do, even if I'm sometimes absent for a long time; Kalina Michalska and Jonas Olofsson, for interesting discussions, help with fMRI and friendship; James Howard: thank you for teaching me fMRI, developing with me one of my studies, and for being a good friend; Joanna Wu: When I first met you, you were the "significance maker" of our study. From then on, you have been an invaluable friend and also teacher (cooking, crochet, pottery, etc.); Neha Mehta: we met over a poor dying bee at the bus after the NUIN retreat, and since then you have been a great friend and a constant supply of positive energy and support.; Thanks also to David Spjut, for doing the piece of artwork that ended up being the cover of my thesis, even though he didn't personally know me; and thanks to Jos Zwanikken, for thinking about it and asking his friend David to do the artwork, and for good discussions, moments, and support. Thanks to Mário Neto, for introducing me to an amazing set of people, for advising me on thesis-writing, and for friendship: I am very glad that you ended up staying much longer in Chicago!; Special thanks to Laetitia Chauve: thank you for your company at coffee shops when writing my thesis; the great moments and trips we spent together; for your constant support and friendship; and in general for always being there; Thanks to Irina Franco, for the constant supply of positive energy, support and friendship; Thanks to Kai Orton: for the comments on my thesis, the uplifting kai-hugs, and for your support and friendliness; Thanks to Bahar Sharafi and Marije De Bruin, for your enthusiasm, support and friendship here at RIC. In general, thank you to all my friends here at Chicago, for amazing moments, celebrations, conversations and friendship that made living in Chicago a great and unforgettable experience.

My friends in Portugal and other parts of the world. I would like to thank: Raquel Antunes: thank you for the comments on my thesis, your support even if oceans apart (glad they invented gchat;)) and in general thanks for being an amazing “mana”; Candida Shinn, for your strength, care and kindness; and Barbara Mauricio, for always being so nice and greeting me with such enthusiasm when I come to Portugal. Thanks to Paula Silva and Anabela Maia, for help in the thesis and a friendship that started when we were small children and still goes on. Thanks to my master thesis colleagues and friends, for making math a much more exciting subject and for being always welcoming: Telma Correia, Catarina Miranda, Teresa Sampaio, Laura Wise, Zé Leal and Ricardo São João. Thanks also to Barbara Vreede and Nuno Sepúlveda, for helping me with my thesis and for friendship. A special thanks also my non-scientist friends, namely Marta Alarcão, Teresa Pessoa, Ana Rita Correia, Ana David, Sara Brasileiro, Alexandre Sousa and Bernardo Santos: thank you for your love and support even when I went to live so far away and wasn’t so present anymore. Thank you all for your amazing friendship.

Finally, I would like to thank my family: without you, nothing of this would ever be possible. I would like to thank: my cousin and “irmona” Alexandra Vilares, for our conversations that developed in me a passion for psychology and friendship; Oma, Margaret Wardenaar-Donga, Rachel Wardenaar, Piet Donga and Stefan Donga, for wanting to fly all the way from the Netherlands just to come to my PhD defense and support me; Pedro, Isaltina and Eduardo Vilares, Annelies and Oscar Donga, Vincent Pietersma, Stijn and Rene Wardenaar. Thank you all for being such a great family. I would specially like to thank my mother, Gezina Donga, my father, Manuel Vilares, and my brother, Ricardo Vilares: you are always there for me, it is amazing how much love you give me, every day. I am very fortunate to have you and I thank you very very much. This thesis is for you.

## LIST OF PUBLICATIONS INCLUDED IN THE THESIS

This thesis is based on the following papers:

**Chapter 2:**

Vilares I, Howard JD, Fernandes HL, Gottfried J and Kording KP (2012). Differential representations of prior and likelihood uncertainty in the human brain. Current Biology, 22(18): 1641-1648.

**Chapter 3:**

Vilares I and Kording KP (*under review*). Decision-making under uncertainty in Parkinson's Disease.

**Chapter 4:**

Vilares I, Dam G and Kording KP (2011). Trust games: Are money and effort equivalent? PLos ONE, 6(2): e17113.

**LIST OF PUBLICATIONS NOT INCLUDED IN THE THESIS**

Stevenson IH, Fernandes HL, Vilares I, Wei K, and Kording KP (2009). Bayesian integration and non-linear feedback control in a full-body motor task. PLoS Computational Biology, 5(12): e1000629.

Vilares I and Kording KP (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Annals of the New York Academy of Sciences, 1224: 22-39.
(although the general introduction included some modified parts of it)

Mhatre P, Vilares I, Stibb S, Albert M, Pickering L, Marciniak C, Kording KP, Toledo S (2013). Wii Fit Balance Board Playing Improves Balance and Gait in Parkinson Disease. Physical Medicine and Rehabilitation, 5 (9): 769-777

Vilares I, Stevenson I, Pavone L, Wei K, Gaebler-Spira D, Kording KP (*submitted*). Wiiagnostics: Activity during gameplay as a diagnostic tool.

**SUMMARY**

Uncertainty pervades most of the events and decisions that we face every day. Uncertainty exists in previously acquired knowledge (prior) and on what our senses currently tell us (likelihood). Moreover, uncertainty exists in both non-social and social settings (social uncertainty). Understanding how the brain responds to and uses information about uncertainty thus seems crucial if we want to understand decision-making.

This thesis sought to understand uncertainty at different levels. We started by asking which areas of the brain are involved in representing and dealing with prior and likelihood uncertainty. We developed a decision-making task in which both prior and likelihood uncertainty were varied and had subjects perform the task while we scanned their brains using functional magnetic resonance imaging (fMRI). We found that the brain represented both types of uncertainty, but they had very distinct representations: likelihood uncertainty activated brain areas along the early stages of the visuomotor pathway, while prior uncertainty was associated with increased activations in specialized brain areas outside this pathway, namely putamen, amygdala, insula and orbitofrontal cortex. Furthermore, activity from the putamen also correlated with subjects' tendencies to sense and attend to current versus prior information. Our results thus suggest different pathways through which prior and likelihood uncertainty are represented in the human brain, offering insights into the neural pathways that allow humans to make decisions.

Based on the obtained results, we next aimed to understand the specific role putamen activity had in decision-making under uncertainty. Furthermore, because dopamine is one of the main neurotransmitters of the putamen and has previously been implicated in decision-making under uncertainty, we also sought to understand the influence dopamine plays in the decision-making process. For that purpose we recruited patients with

Parkinson's disease (PD), in whom the putamen activity is compromised due to a lack of dopamine, to perform the same task, and compared them with age-matched controls. We hypothesized that if dopaminergic activity from the putamen is causally involved in some aspect of decision-making under uncertainty, this will then show in the way PD patients perform the task. Moreover, we hypothesized that these differences would be particularly salient if PD patients performed the task after they were off dopaminergic replacement medication overnight (*off-state*), compared with if they just took their medication 1 hour before (*on-state*). We found that some aspects of decision-making under uncertainty were conserved in PD patients: both groups could learn prior distributions and use prior and likelihood information. However, PD patients, particularly in the *off-state*, were impaired at reacting to differences in likelihood uncertainty. Together, our results indicate that dopaminergic activity from the putamen has a crucial role in the processing of uncertainty in the current sensory stimulus (likelihood uncertainty), potentially through increased attention towards it.

Finally, we asked how people deal with uncertainty stemming from social interactions. Specifically, we aimed at understanding if trust, which can be defined as a positive expectation in the face of social uncertainty, would vary depending on whether the situation was framed in a monetary or movement effort way. We developed a trust game based on movement effort and compared the results with those of a computational equivalent monetary trust game. We found no difference in trust between both conditions. Moreover, there was a high positive correlation in subjects' behavior across conditions. These results suggest that the way people deal with social uncertainty may be a character trait: subjects that trust more in monetary settings behave similarly during exchanges of movement effort.

In summary, the work presented in this thesis advances the understanding of how people deal with different kinds of uncertainty, suggesting precise roles for specific brain areas and for dopamine in decision-making under uncertainty.

**RESUMO**

A incerteza pervade a maioria dos acontecimentos e decisões que temos de tomar no dia-a-dia. Existe incerteza no conhecimento previamente obtido (distribuição *a priori*) e no que os nossos sentidos nos dizem correntemente (verossimilhança). Além disso, existe incerteza não só em contextos não-sociais mas também em contextos sociais (incerteza social). Uma melhor compreensão de como o cérebro humano responde e usa informação sobre incerteza parece portanto crucial se quisermos entender melhor como são tomadas decisões.

Esta tese teve como objectivo tentar perceber a incerteza a vários níveis. Começámos por tentar perceber que áreas do cérebro estão envolvidas na representação e uso da incerteza relacionada com a informação *a priori* e com a verosimilhança. Para isso, desenvolvemos um teste em que variámos tanto a incerteza relacionada com a distribuição *a priori* como a relacionada com a informação corrente (verossimilança), e pedimos a voluntários para fazer o teste enquanto os submetiamos a um exame de ressonância magnética funcional (fMRI). Os nossos resultados indicaram que o cérebro representa ambos os tipos de incerteza, mas que estas têm representações muito distintas: a incerteza na informação corrente activou regiões cerebrais ao longo da via visuo-motora, enquanto que a incerteza relacionada com a distribuição *a priori* estava associada com maior activação em regiões cerebrais especializadas fora desta via, nomeadamente o putamen, a amígdala, a ínsula e o cortex prefrontal. Além disso, a actividade do putamen estava também correlacionada com tendências individuais para notar e reagir a informação corrente versus *a priori*. Os nossos resultados sugerem vias diferentes pelas quais a incerteza associada à informação *a priori* e à informação corrente são representadas no cérebro humano, apresentando os possíveis mecanismos neuronais que possibilitam ao ser humano tomar decisões com incerteza associada.

Face aos resultados obtidos, procurámos de seguida perceber qual a função específica que a actividade neuronal do putamen tem na tomada de decisões com incerteza associada. Mais ainda, quisemos também saber se a dopamina estaria envolvida no processo, uma vez que é um dos neurotransmissores mais importantes do putamen e foi já implicada previamente em tomadas de decisão com incerteza. Pedimos a pacientes com a doença de Parkinson, nos quais a actividade neuronal do putamen está afectada devido à falta de dopamina, para fazer o mesmo teste, e comparámos o seu comportamento com um grupo controle da mesma idade. A nossa hipótese era que, se a actividade dopaminérgica do putamen estivesse causalmente envolvida nalgum dos aspectos de tomada de decisão com incerteza associada, então isso iria ser visível na forma como os doentes de Parkinson fazem o teste. Além disso, pusémos a hipótese de que essas diferenças seriam particularmente salientes se os doentes de Parkinson tivessem feito o teste depois de terem estado sem tomar medicação dopaminérgica durante várias horas (*off-state*) em comparação com a situação em que tivessem tomado a medicação 1 hora antes (*on-state*). Verificámos que vários aspectos da tomada de decisão com incerteza associada estavam intactos em doentes de Parkinson: ambos os grupos (pacientes e controlos) conseguiram aprender a distribuição *a priori* e usar informação tanto da distribuição *a priori* como da verossimilhança. Contudo, os doentes de Parkinson, especialmente se estivessem em *off-state*, tinham problemas em reagir a diferenças na incerteza associada à verossimilhança. De um modo geral, os nossos resultados indicam que a actividade dopaminérgica do putamen tem uma função crucial no processamento da incerteza associada ao estímulo sensorial corrente (incerteza na verossimilhança), potencialmente através do aumento da atenção dirigida a este.

Por último, perguntámos como é que as pessoas lidam com a incerteza proveniente de interacções sociais. Mais especificamente, queríamos perceber se a confiança noutras pessoas, que pode ser definida

como uma expectativa positiva face à incerteza associada à interacção social, iria variar dependendo se a troca social fosse apresentada como uma troca monetária ou de esforço físico. Desenvolvemos para isso um jogo de confiança baseado em esforço físico e comparámos os resultados com um jogo de confiança monetário computacionalmente equivalente. Não foram encontradas diferenças em termos de confiança entre as diferentes condições. Além disso, verificámos que o comportamento dos voluntários numa condição estava fortemente e positivamente correlacionado com o seu comportamento na outra condição. Estes resultados sugerem que a forma como as pessoas lidam com incerteza social pode ser uma característica da sua personalidade: pessoas que confiam mais noutras pessoas numa situação monetária também confiam mais numa situação em que é requerido esforço físico.

Em resumo, o trabalho apresentado nesta tese avança o conhecimento de como é que as pessoas lidam com diferentes tipos de incerteza, sugerindo funções específicas para certas regiões cerebrais e para a dopamina no processo de tomada de decisões com incerteza associada.

**CONTENTS**

APOIO FINANCEIRO/FINANCIAL SUPPORT ..... v
ACKNOWLEDGEMENTS ..... vii
LIST OF PUBLICATIONS INCLUDED IN THE THESIS ..... xiii
LIST OF PUBLICATIONS NOT INCLUDED IN THE THESIS ..... xv
SUMMARY ..... xvii
RESUMO ..... xix
CONTENTS ..... xxiii
1. GENERAL INTRODUCTION ..... 3
1.1 Uncertainty is everywhere ..... 3
1.2 How to deal with uncertainty? Bayes theory ..... 3
1.3 Humans can combine current sensory information with prior knowledge ..... 9
1.4 Bayesian decision-making ..... 11
1.5 Theories on neural representations of uncertainty ..... 13
1.6 Social uncertainty ..... 18
1.6.1 Trust ..... 18
1.6.2 Game theory ..... 19
1.6.3 Trust games ..... 21
1.7 This thesis: objectives and outline ..... 22
1.8 References ..... 24
2. DIFFERENTIAL REPRESENTATIONS OF PRIOR AND LIKELIHOOD UNCERTAINTY IN THE HUMAN BRAIN ..... 37
2.1 Summary ..... 37
2.2 Introduction ..... 39
2.3 Materials and methods ..... 43
2.4 Results ..... 55
2.4.1 Main results ..... 56

2.4.2 Additional results ..... 68
2.5 Discussion ..... 79
2.6 Acknowledgements ..... 83
2.7 References ..... 83
3. DECISION-MAKING UNDER UNCERTAINTY IN PARKINSON'S DISEASE ..... 91
3.1 Summary ..... 91
3.2 Introduction ..... 93
3.3 Materials and methods ..... 97
3.4 Results ..... 105
3.5 Discussion ..... 112
3.6 Acknowledgements ..... 117
3.7 References ..... 117
4. TRUST GAMES: ARE MONEY AND EFFORT EQUIVALENT? ..... 127
4.1 Summary ..... 127
4.2 Introduction ..... 129
4.3 Materials and methods ..... 133
4.4 Results ..... 137
4.5 Discussion ..... 151
4.6 Acknowledgements ..... 154
4.7 References ..... 154
5. FINAL DISCUSSION ..... 161
5.1 This thesis: brief overview and conclusions ..... 161
5.1.1 Neural correlates of Bayesian statistics ..... 161
5.1.2 Role of dopamine in decision-making under uncertainty ..... 163
5.1.3 Dealing with social uncertainty as a potential character trait ..... 164
5.2 Limitations of the work and future research ..... 166
5.3 Potential implications ..... 168
5.4 Conclusion ..... 170

5.5 References ..... 172
6. APPENDIX ..... 179

# 1. GENERAL INTRODUCTION

*“The truth would be literally nothing but the shadows of the images.”*

*Plato*

# 1. GENERAL INTRODUCTION

## 1.1 Uncertainty is everywhere

Uncertainty is everywhere. The information that we get from the world at any point of time is uncertain, because there is inherent noise in our senses. Additionally, any information that we have must be somewhat uncertain, since we only observe incomplete portions of the world, or shadows of the reality, as beautifully illustrated by Plato’s allegory of the cave (Plato, 360 B.C.). Nevertheless, every day we have to make decisions that have some associated uncertainty, ranging from small unconscious decisions (e.g. where to step next) to potentially life-changing situations (e.g. what career should I choose? should I marry this person?). Therefore it may be argued that a central purpose of the nervous system is to estimate the state of the world and make decisions based on uncertain information (Helmholtz, 1856, Smith, 2001).

## 1.2 How to deal with uncertainty? Bayes theory

How does, then, the brain deal with uncertainty? A related question that we can ask is how should the brain deal with it? This second question differs from the first one in the sense that it not only aims at describing how a certain problem is solved, but it tries to find what would be the optimal solution for it. This sort of reasoning drives a line of research called *normative models*, which start with an idea of what the objective of a given system could be and then derive what would be the optimal solution to arrive at that objective. The model predictions are then usually compared with the way the system actually behaves or is organized (Kording, 2007). In

this way, normative models can test hypotheses about the potential purpose of parts of the nervous system.

The best way to deal with uncertain information is given by *Bayes theory* (Glimcher, 2003). Bayesian statistics gives a systematic way of calculating optimal estimates based on noisy or uncertain data. Bayesian statistics models of the nervous system start with the idea that the nervous system needs to estimate variables in the world that are relevant ($X_{true}$) based on sensed information ($X_{sensed}$), typically coming from our sensory systems (e.g. audition, vision, olfaction). Bayes rule (Bayes, 1764) then allows calculating how probable it is that a certain hypothesized x is the true x ($x_{true}$) given the sensed information $x_{sensed}$:

$$p(x_{true}|x_{sensed}) = \frac{p(x_{sensed}|x_{true})p(x_{true})}{p(x_{sensed})} \quad (1)$$

For example, consider that you are a goalkeeper in a soccer match and you have a player from the other team shooting the ball towards your goal (see the cover of this thesis). In order to try to prevent the other team from scoring a goal, you have to estimate the position of the ball ($x_{true}$ = "position of the ball") given what your senses are telling you now ($x_{sensed}$ = "sensed position"). In our example, let's say that you are trying to guess the probability that the ball is being kicked to the right side of the goal given what you are seeing at the moment (i.e. you are trying to estimate $p(x_{true} \mid x_{sensed})$). You can then combine the probability of observing the ball at that particular location if the ball was being kicked to the right ($p(x_{sensed} / x_{true})$) with the previous knowledge that you have about where this specific player generally kicks the ball ($p(x_{true})$). Mathematically, the equation above can be seen as a way of updating the previous belief about the world, or **prior** ($p(x_{true})$) by the *current sensory information*, or **likelihood** ($p(x_{sensed} / x_{true})$),

divided by a normalizing constant ($p(x_{sensed})$), in order to arrive to a **posterior** probability ($p(x_{true} \mid x_{sensed})$). You can then calculate a posterior probability for each of your hypotheses (e.g. the ball is going to the right or to the left), choose the one with the highest probability, and behave accordingly (e.g. throw yourself to the right or the left to try to defend the goal)[1].

If we assume that both the prior and current sensory information follow a normal distribution, then Equation 1 can be written as (Kording and Wolpert, 2004):

$$p(x_{true}|x_{sensed}) = \frac{\frac{1}{\sigma_{sensed}\sqrt{2\pi}} e^{-\frac{(x_{true}-x_{sensed})^2}{2\sigma_{sensed}^2}} \frac{1}{\sigma_{prior}\sqrt{2\pi}} e^{-\frac{(x_{true}-x_{prior})^2}{2\sigma_{prior}^2}}}{p(x_{sensed})}$$

In order to get the optimal estimate we can find the maximum by differentiating and equaling to zero (Kording and Wolpert, 2004). The resulting optimal estimate is basically a weighted sum of the mean of the prior and the current sensory information (in our example it would be the perceived position of the ball):

$$x_{est} = \frac{\sigma_{sensed}^2}{\sigma_{sensed}^2+\sigma_{prior}^2}\mu_{prior} + \frac{\sigma_{prior}^2}{\sigma_{sensed}^2+\sigma_{prior}^2}\mu_{sensed} \qquad (2)$$

[1] This if you are using a maximum-a-posteriori (MAP) approach.

where $x_{est}$ is the optimal estimate (e.g. the optimal estimate of the position of the ball), $\sigma^2_{sensed}$ and $\sigma^2_{prior}$ are the variances of the likelihood and of the prior, respectively, and $\mu_{sensed}$ and $\mu_{prior}$ are the respective means.

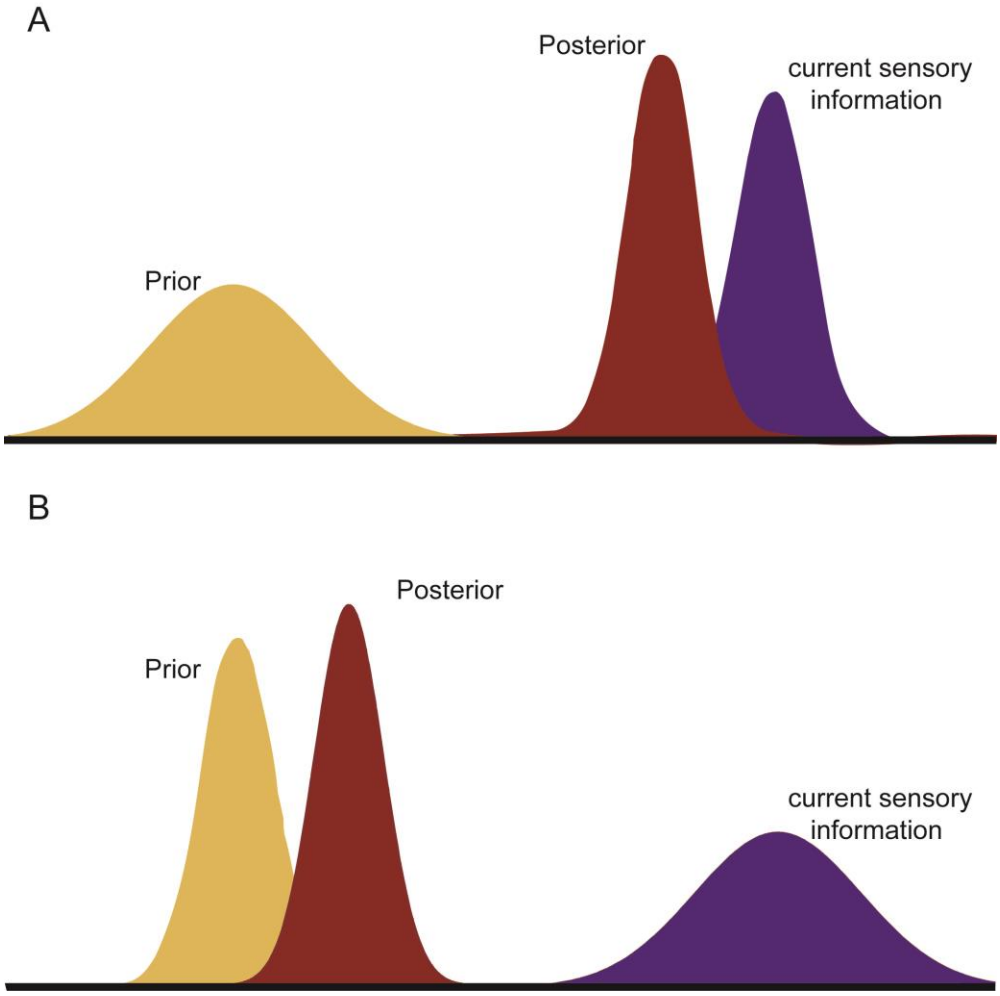


**Figure 1.1 –Schematic illustration of the combination of the prior distribution with the distribution of the possible currently sensed information in order to arrive to a posterior.** A) The currently sensed information (in purple) has a smaller variance (uncertainty) than the prior (in yellow), and hence the mean of the resulting posterior distribution (in purple) will be closer to the mean of the currently sensed information. B) The prior is narrower than the currently sensed information, and so the mean of the posterior is closer to the mean of the prior.

Notice that in Equation 2 the information about the prior and the currently observed sensory information (likelihood) is weighted according to the respective relative reliabilities. For example, if the prior information is much more uncertain than the likelihood, i.e. if $\sigma^2_{prior} >> \sigma^2_{sensed}$ (e.g. if we never interacted with that player before and we don't know where he generally shoots to), then $\sigma^2_{sensed} / (\sigma^2_{sensed} + \sigma^2_{prior})$ will be closer to zero and $\sigma^2_{prior} / (\sigma^2_{sensed} + \sigma^2_{prior})$ closer to one, and consequently the prior will have little influence on the final estimate, which will be much closer to the currently observed sensory information. Graphically (see Figure 1.1-A), we can see how the posterior distribution will be closer to the distribution of the current sensory information. Alternatively, if the current sensory information is noisier than the prior information (e.g. in a foggy day), then the final estimate will be more biased towards the mean of the prior (Figure 1.1-B).

Equation 2 has been used to experimentally calculate the weight given to current versus prior information and also to infer how uncertain someone feels about each of these pieces of information (Kording and Wolpert, 2004, 2006). Assuming that the mean of the prior remains constant, one can give subjects sensory stimuli with different observed means. This allows one to create a linear regression of the subjects' estimates ($X_{est}$) as a function of the observed sensory mean ($\mu_{sensed}$):

$$x_{est} = \underbrace{\frac{\sigma^2_{sensed}}{\sigma^2_{sensed}+\sigma^2_{prior}}\mu_{prior}}_{\beta_0} + \underbrace{\frac{\sigma^2_{prior}}{\sigma^2_{sensed}+\sigma^2_{prior}}\mu_{sensed}}_{\beta_1}$$

The intercept of the linear regression ($\beta_0$) would be equal to $\sigma^2_{sensed} / (\sigma^2_{sensed} + \sigma^2_{prior})\ \mu_{prior}$. The slope of this linear regression ($\beta_1$) characterizes how much

the subject is weighting the current sensory information, and, if people perform according to the optimum prescribed by Bayesian statistics, its value should be equal to the perceived $\sigma^2_{prior}/(\sigma^2_{sensed}+\sigma^2_{prior})$. A slope of zero indicates that subjects do not take into account current information, and a slope of one indicates that subjects only use current information. A slope between zero and one indicates that subjects are using information from both prior and current information, and the larger the slope the more they rely on the current sensory information and less on the prior. The slope of this linear regression, which we will call the *Bayesian slope*, tells us then how much the subject weights current relative to prior information.

One can experimentally vary the uncertainty (variance) of the prior and/or the current information and calculate how much it changes subjects' Bayesian slope (Kording and Wolpert, 2004, 2006, Tassinari et al., 2006, Beierholm et al., 2009, Berniker et al., 2010). These values can then be compared with what would be expected from Bayesian statistics: qualitatively, if people are performing according to Bayesian statistics, then this slope should increase with increased uncertainty in the prior and decreased uncertainty in the current information. Quantitatively, it should be near $\sigma^2_{prior}/(\sigma^2_{sensed}+\sigma^2_{prior})$. Furthermore, if one assumes that subjects use a Bayesian strategy, it is also possible to infer subjects' perceived uncertainty in prior and current sensory information (Kording and Wolpert, 2004).

Another interesting property of Bayesian estimation is that the variance of the optimal estimate is equal to $\sigma^2_{sensed}\,\sigma^2_{prior}/(\sigma^2_{sensed}+\sigma^2_{prior})$, and hence its variance is lower[2] than the variance of each piece of information (prior and current sensory observation) alone, i.e. $\sigma^2_{sensed}\sigma^2_{prior}/(\sigma^2_{sensed}+\sigma^2_{prior})\leq\sigma^2_{sensed}$ and $\sigma^2_{sensed}\sigma^2_{prior}/(\sigma^2_{sensed}+\sigma^2_{prior})\leq\sigma^2_{prior}$. This property of lower uncertainty in the final estimate is one of the crucial

[2] Being only equal if one piece of information has no uncertainty, i.e. if $\sigma^2_{prior}=0$ or $\sigma^2_{sensed}=0$

advantages of behavior that combines different pieces of knowledge in a way predicted by Bayesian models.

Models using Bayes rule have been used to explain many results in perception, action, neural coding and cognition (Ernst and Banks, 2002, Kording and Wolpert, 2006, Ma et al., 2006, Ma et al., 2009, Maia, 2009). Bayesian models that have been used in these contexts have many different forms. The differences between these models derive from distinct assumptions about the variables in the world and the way they relate to one another. Each model is then the unique consequence of one set of assumptions about the world. However, all these Bayesian models share the same basic principle that different pieces of information can be combined in order to estimate the relevant variables (Vilares and Kording, 2011).

## 1.3 Humans can combine current sensory information with prior knowledge

There are many examples of the effect of prior knowledge in perception. For example, in Figure 1.2 we can see one concave groove and two convex bumps, but if we rotate the paper by 180 degrees their perceptual depth shifts (we see two grooves and one bump instead). This occurs because people have the prior assumption that light should come from above. This effect of prior knowledge is also beautifully illustrated in the Checker-shadow illusion, where the prior assumption of a light source that casts the observed shadow makes the rectangles A and B appear to have different brightness (Adelson, 1995). These sensory biases can then be explained with the incorporation of prior information on the final sensory perception.

Experimental studies have shown that priors can indeed be learned over time (Berniker et al., 2010) and that they are independent of the current sensory information (Beierholm et al., 2009). A diverse set of studies has

also shown that people often combine previously acquired knowledge (prior) with current sensory information (likelihood) in a way that is close to the optimum predicted by Bayesian statistics (but see Beck et al., 2012). For example, when performing arm-reaching tasks (Kording and Wolpert, 2004, Brouwer and Knill, 2009), pointing tasks (Tassinari et al., 2006) or even timing tasks (Miyazaki et al., 2005), people take into account both prior and likelihood information and, moreover, they do so in a way compatible with Bayesian statistics, giving more weight to the more reliable cue. In other words, they rely more on the prior when likelihood is a poor source of information and vice versa (Kording and Wolpert, 2004, Miyazaki et al., 2005, Tassinari et al., 2006, Brouwer and Knill, 2009, Gerardin et al., 2010).

If people are taking into account the uncertainty in both the prior and the likelihood, this means that the brain likely somehow represents these uncertainties. How does the nervous system then represent prior and likelihood and their associated uncertainties? Currently there is relatively little known about how the brain represents uncertainty (although some theories have been proposed, see subsection 1.5). When it comes to movement, it has been found that the dorsal premotor cortex and the primary motor cortex encode multiple potential reaching directions, indicating a potential representation of priors (Riehle and Requin, 1989, Bastian et al., 1998, Cisek and Kalaska, 2002, Bastian et al., 2003, Cisek and Kalaska, 2005). However, we clearly do not yet understand how the nervous system integrates priors and likelihoods, or where the respective uncertainties are represented.

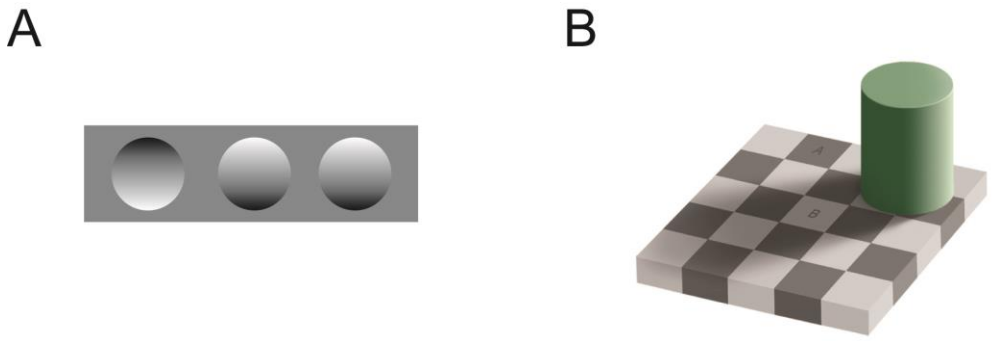


**Figure 1.2 - Examples of the effect of prior knowledge in current sensory perception.** A) Example of a visual illusion. Here people generally see one groove (left) and two bumps, but if the paper is rotated 180° then two grooves and one bump are perceived. B) Checker-shadow illusion (Adelson, 1995). In this visual illusion, the rectangle A appears to be darker than B, while in reality they have the same color.

## 1.4 Bayesian decision-making

So far we have seen how prior and current sensory information can be combined, but that is only the first step in decision-making systems. In our example, after estimating the probabilities associated for each potential ball's location, we have to choose how to try to defend it. Depending on the way we do that, we may incur a cost: For example if a ball is estimated to go in the corner of the goal, then it would be good to jump to it. If, on the other hand, it is estimated to be kicked in the center of the goal, then one should stay in place, and jumping to the corner may be costly. In the field of economics, the costs are generally described as negative utility, and the rewards as positive utility, where utility measures the subjective value of any possible situation. Decision theory deals with this problem of choosing the right action given uncertainty, generally by calculating the action that maximizes expected utility (Bentham, 1780). To make good decisions we need to combine our uncertain knowledge of the world with the potential rewards and costs we may encounter.

Sensorimotor research has shown that when human subjects are doing a movement task they are able to estimate their motor uncertainties (Christopoulos and Schrater, 2009), take into account both rewards and penalties associated with it, and aim their movements in a way that maximizes expected utility (Maloney et al., 2006). This is in contrast to many high-level economics tasks where human subjects exhibit a wide range of deviations from optimality (Kahneman and Tversky, 1979, Ariely, 2008).

To maximize expected utility our brain has to represent not only the uncertainty associated to each action but also its respective reward or cost value. The neural representation of these variables has been the focus of the emerging field of Neuroeconomics. Neuroeconomics tries to understand the neural processes that occur during decision-making within the framework of Bayesian decision theory (Glimcher, 2003, Wolpert, 2007, Beck et al., 2008). Responses to reward value and reward probability have been identified in neurons in the orbitofrontal cortex, striatum, amygdala and dopamine neurons of the midbrain (Cromwell and Schultz, 2003, Fiorillo et al., 2003, Gottfried et al., 2003, Tobler et al., 2005, Padoa-Schioppa and Assad, 2006, Paton et al., 2006).

In recent years, research in neuroeconomics has been shifting focus from the mere coding of expected value and magnitude of reward to neural representations of reward uncertainty, both in the form of risk and of ambiguity. Uncertainty in reward has been hypothesized to be represented in some of the brain areas typically associated with reward coding, such as dopamine neurons (Fiorillo et al., 2003), the amygdala (Delazer et al., 2010), the orbitofrontal cortex and the striatum (Preuschoff et al., 2006, Tobler et al., 2007) . However, it has been noted that neuronal activations related to uncertainty in reward (more specifically, to risk) seem to be segregated spatially and temporally from activations due to expected reward, with activations due to risk occurring later than the immediate activations due to expected reward (Preuschoff et al., 2006). Besides these

activations in reward-related areas, uncertainty in reward has also been associated with unique activations in the insula (Huettel et al., 2005, Singer et al., 2009) and in the cingulate cortex (McCoy and Platt, 2005, Rushworth and Behrens, 2008). Thus, reward value and risk have been associated with spatially and temporally distinct neural activations in specific brain areas, suggesting that the brain can use both sources of information to estimate expected utility and guide actions.

Research in neuroeconomics has thus given us valuable suggestions about where uncertainty in reward can be represented in the brain. In this research, uncertainty is generally treated as one parameter, which can be considered the posterior uncertainty of a rewarding event. However, it does not differentiate between prior and current sensory uncertainty. In their experiments prior and current sensory uncertainty were perfectly correlated with posterior uncertainty, and were not distinguished as different sources of uncertainty. It is still unclear where prior and current sensory uncertainty are represented, and if they are represented in the same areas associated with uncertainty in reward.

## 1.5 Theories on neural representations of uncertainty

There are many different theories that have been proposed to describe how the brain may represent uncertainty (see Figure 1.3). For example, uncertainty can be encoded in a neuron's tuning curve (Figure 1.3-A:C are examples of how changes in uncertainty could affect tuning curves). A tuning curve represents the average firing rate of the neuron as a function of relevant stimulus parameters (Butts and Goldman, 2006). There are at least two ways by which uncertainty could be encoded in a neuron's tuning curve: either there are specialized neurons that encode uncertainty and nothing else (Figure 1.3-A) or alternatively neurons may have tuning curves

representing some other variable and encoding at the same time the uncertainty associated with that variable (Figure 1.3-B and Figure 1.3-C).

The first theory depicted in Figure 1.3-A is the simplest: there may be a subset of neurons that only encode uncertainty, for example using neuromodulators such as dopamine, acetylcholine or norepinephrine (Yu and Dayan, 2005). This theory appears to have significant experimental support. For example, some experiments, as discussed above, indicate that uncertainty about reward appears to be represented by groups of dopaminergic neurons in the substantia nigra, insula, orbitofrontal cortex, cingulate cortex and amygdala (Fiorillo et al., 2003, Huettel et al., 2005, McCoy and Platt, 2005, Kepecs et al., 2008, Singer et al., 2009). However, even neurons in these areas have clear tuning to other variables. Furthermore, the experiments in support of this theory generally refer to high level uncertainty, such as the uncertainty associated with potential rewards, and not uncertainty that is related to say sensory or motor information, which can be considered more “low level uncertainty” and might be represented in a different way.

A second possibility states that the width of the tuning curves may change with uncertainty (Anderson, 1994) and that the neurons may jointly encode probability distributions (Figure 1.3-B). Such a joint encoding could be plausible given that the visual system exhibits far more neurons than inputs (Van Essen et al., 1992) and the extra neurons could encode probability distributions instead of point estimates. In conditions of high uncertainty a broad set of neurons would be active but exhibit low activity. In contrast, when uncertainty is low, then only few neurons would be active, but these would have high firing rates (see Figure 1.3-B). Support for this theory comes from early visual physiology where spatial frequency tuning curves of neurons in the retina are larger during darkness (when there is more visual uncertainty) than during the day (Barlow et al., 1957).

A third influential theory of the encoding of uncertainty is the so-called *probabilistic population code*, or PPC (Figure 1.3-C). This theory starts with the observation that the Poisson-like[3] firing observed for most neurons automatically implies uncertainty about the stimulus that is driving a neuron (Ma et al., 2006). In this way, neurons transmit the stimulus information while at the same time jointly transmitting the uncertainty associated with that stimulus. Specifically, the standard versions of this theory predict that increased firing rates of neurons imply decreased levels of uncertainty. Some data in support of this theory comes from studies on cue combination (Gu et al., 2008, Morgan et al., 2008). More support comes from the general finding that early visual activity is higher when contrast is higher and thus uncertainty is lower (Shapley et al., 1981, Carandini and Heeger, 1994, Cheng et al., 1994). However, there is little experimental support for Poisson-like variability, and more advanced population decoding studies are needed.

Another theory suggests that while the tuning curves stay the same the relative timing of signals may change (Figure 1.3-D)(Deneve, 2008, Huan and Rao, 2010). If uncertainty is low then neurons would have high firing rates and fire rapidly when a stimulus is given, but would more quickly stop firing. If, on the other hand, uncertainty is high, then neurons would have lower firing rates but fire for a longer time. In that way, the total number of spikes could be the same, but their relative timing would change. There is some evidence for this theory coming from studies in the area MT that shows differential temporal modulation when animals are more uncertain (Bair and Koch, 1996).

Another theory is the sampling hypothesis (Hinton and Sejnowski, 1983, Hoyer and Hyvärinen, 2003, Fiser et al., 2010, Berkes et al., 2011).

[3] i.e. if the firing of the neuron follows a Poisson-like distribution.

According to this theory, neurons will spike a range of instantaneous firing rates that is narrow over time if the nervous system is certain about a variable and has a wider range if it is less certain (see Figure 1.3-E). Evidence for the sampling hypothesis comes from some recent experiments comparing the statistics of neuronal firing across different situations (Kenet et al., 2003, Fiser et al., 2004). However, no experiments to our knowledge have explicitly changed probabilities and measured the resulting neuronal variability.

Finally, it is possible that uncertainty could be encoded not in the firing properties of neurons but in the connections between them (see Figure 1.3-F), for example in the number and strength of synapses between them (Wu and Amari, 2003). This type of uncertainty coding could for example apply to the representation of uncertainty in priors, as priors are acquired over long periods of time and thus there is a need to store information in a more durable way. Uncertainty, in this case, would thus change the way that neurons interact with one another.

Although all these theories provide interesting suggestions for how the nervous system could represent uncertainty, so far available experimental data does not clearly support one theory over any of the others. More importantly, these theories (portrayed in Figure 1.3) are not mutually exclusive. The nervous system may use any or all of these mechanisms to encode uncertainty at the same time, and use different types of coding for different types of uncertainty. It is thus important to try to acquire more experimental data on how different types of uncertainty may be represented in the brain.

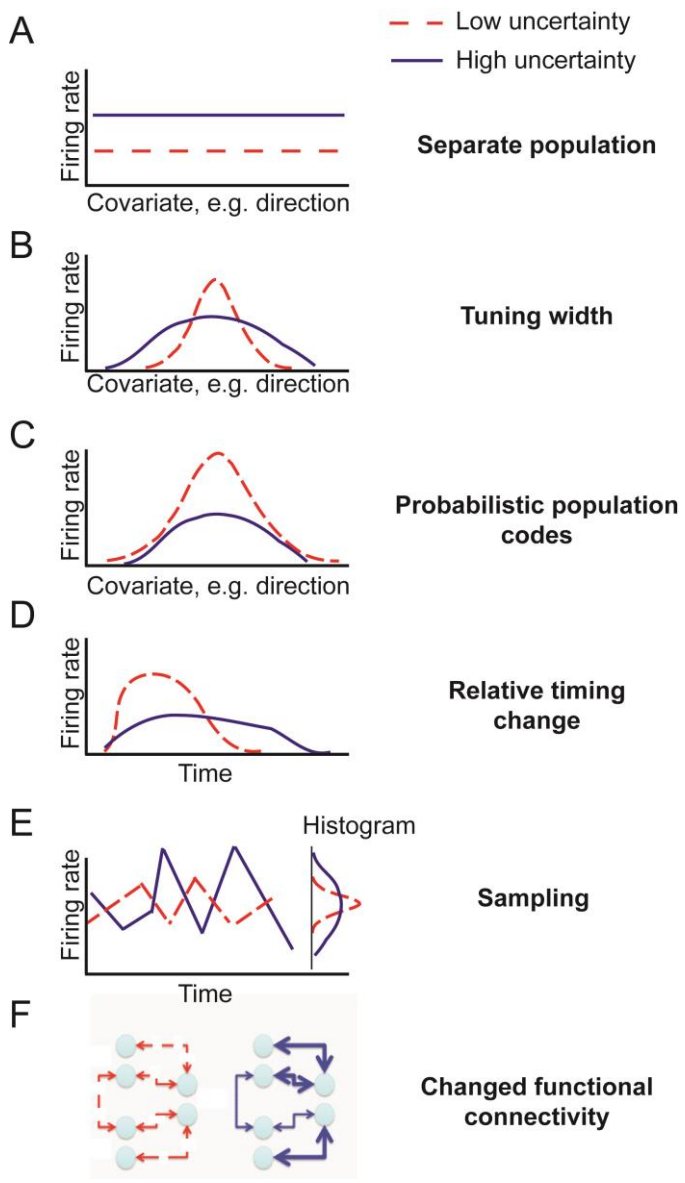


**Figure 1.3 - Possible neural representations of uncertainty.** In red are the putative firing rates (or connections) in a low-uncertainty state and in blue the ones occurring in a high-uncertainty state. Panels A trough F represent different theories that have been proposed on how the brain could be representing uncertainty.

## 1.6 Social uncertainty

One big source of uncertainty in decision-making is other people's behavior. In our example, after estimating the probabilities associated with each potential ball's location, understanding the available courses of action (e.g. going left or right) and incorporating the likely rewards and costs associated with each of them, one can choose the action that maximizes expected utility (Bentham, 1780). However, when interacting with other people, the outcomes of our decisions/actions will also depend on the behavior of others, which is to a certain extent unknown for us. For example, another player can place himself between you and the ball, preventing you from reaching it. Or the player from the other team that is kicking the ball purposefully pretends to be throwing the ball to the right, while ending up throwing to the left. If we take into account other people's behavior we can formulate more accurate predictions of the outcomes of our actions (Osborne and Rubinstein, 1994), and in this way make decisions that lead to better outcomes.

### 1.6.1 Trust

As we saw with non-social events, making predictions of the behavior of other people implies combining prior knowledge, or expectations, with current observable information. Current observable information, however, is uncertain, either because not enough evidence could be gathered and/or because this evidence could have multiple interpretations (Hinde and Groebel, 1991). This is especially true in the case of one-shot interactions, in which no *a priori* information is known about the person we are interacting with (Osborne and Rubinstein, 1994). One has then to incorporate prior beliefs or expectations when making the predictions. Trust can be defined as having the prior belief that the person we are interacting with will promote our best interests and will not betray us (Simpson, 2007). Trust is then a

positive expectation in the face of uncertainty emerging from social relations (Guseva and Rona-Tas, 2001).

Trust facilitates cooperation with others and exists to some degree in all human interaction, being the basis of a productive and healthy society (Camerer, 2003, Krueger et al., 2007). It facilitates organization in both permanent and temporary work groups and is associated with higher job satisfaction, lower labor cost and larger profits (Meyerson et al., 1996, Knack and Keefer, 1997, Gambetta, 2000, Chami and Fullenkamp, 2001). It is also the basis of fulfilling and secure human relationships (Simpson, 2007, Luchies et al., 2013). Trust can be seen in a diverse array of situations, and over a variety of things: people entrust money to insurance companies, hoping that they will help them in return if the time comes; they trust physical effort, for example by helping a friend move; and they even trust their own lives to complete strangers, every time they step into a bus or plane that is not driven or piloted by them. Furthermore, while traditionally people would mainly live and interact in small communities, and so long-term repeated interactions would allow trust to be built over time, in the current global market these long-term repeated exchanges between relatives or neighbors are being slowly replaced by one-time interactions between anonymous partners (Nowak and Sigmund, 2005). A better understanding of how trust operates in different contexts in one-shot interactions may then be of special relevance to the current economy.

### 1.6.2 Game theory

Game theory provides a collection of rigorous models that were developed to understand and explain situations in which two people (decision-makers) must interact, and in which each person choice behavior will affect the outcome of other player (Osborne and Rubinstein, 1994, Camerer, 2003, Sanfey, 2007). It offers a rich source of both behavioral tasks and data, in

addition to well-specified models for the investigation of social exchange (Sanfey, 2007). Although it was first used in economics, it has now been employed in a wide variety of fields, including sociology, evolutionary biology and the emerging field of neuroeconomics (Osborne and Rubinstein, 1994, Camerer, 2003, Sanfey, 2007). Given its mathematic basis, it allows for quantitative predictions of how idealized players should behave, and deviations from this "optimal behavior" can give insights into people's internal motivations, cognitive limitations and prior expectations (Camerer, 2003).

In game theory, the basic entity is a "player", and a "game" contains a specification of the actions each player has available to them, the information players have (complete or incomplete), the order in which they choose their actions (e.g. all at the same time or not) and the possible outcomes ("utilities") associated with each set of actions (Osborne and Rubinstein, 1994, Camerer, 2003). For each game it is possible to calculate the Nash Equilibrium, i.e. the set of actions from which no player can benefit from changing unilaterally his or hers action (Osborne and Rubinstein, 1994, Camerer, 2003). The Nash equilibrium tells us what a player should do, i.e. what is the optimal or best response given the behavior of the other players, assuming that the player is "rational" (just interested in maximizing their own payoff). However, in practice many times people deviate from what would be expected under the Nash equilibrium, being less selfish and strategic that the game theoretical models predict (Sanfey, 2007). The way people deviate from the Nash equilibrium can then, as referred, give valuable information about people's internal motivations and prior expectations, and the formal modeling approach and well-characterized tasks given by Game Theory provide a very useful foundation for the study of decisions in a social context (Sanfey, 2007).

### 1.6.3 Trust games

Game theory offers a simplified way to define trust and measure it, by using the *trust game* depicted in Figure 1.4. In a trust game, one person (the *investor*, or *trustor*) receives a given amount of money, and can choose to invest some part of it. The amount invested is then multiplied by some factor higher than 1 (symbolizing a return on social investment) and given to the other player (the *trustee*). The trustee can then decide how much of this increased amount to give back to the trustor, keeping the rest (Berg et al., 1995, Camerer, 2003). From a rational choice point of view, if the trustor does invest money, this represents a belief that the expected return will be positive. The amount of money invested is then a measure of the *trust* deposited on the trustee, and the money given back by the trustee is a measure of *reciprocity*.

Many studies have used trust games and the results have been contrary to what would be expected under the assumption of purely self-interested individuals, who act in order to maximize their own payoff (Camerer and Fehr, 2002). In fact, if the trust game is played only once (one-shot game), the unique Nash equilibrium for rational purely selfish players is for the trustee to give no money back, and so for the trustor (anticipating that) to invest nothing (Camerer, 2003, King-Casas et al., 2005). Nevertheless, studies have consistently shown that people do trust and reciprocate, even at a cost to their gains (Berg et al., 1995, Fehr and Gachter, 2000, Camerer, 2003). However, these results were mainly obtained using exchanges of money, and it is unclear if they are only applicable to monetary transactions, or if this “monetary trust” indeed represents a general positive expectation in the face of social uncertainty (Levitt and List, 2007).

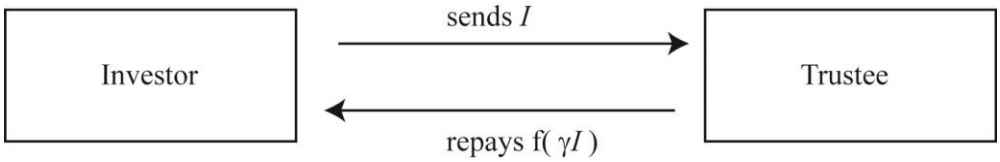


**Figure 1.4 - Schematic representation of a trust game.** The Investor receives money in the beginning of the experiment and can decide, from that money, to send an amount *I* to the trustee, keeping the rest. The amount invested is then multiplied by a factor $\gamma$ (which is higher than 1) and given to the *trustee*. The trustee can then decide which fraction of this increased amount ($\gamma I$) to give back to the trustor, and keeps the rest.

## 1.7 This thesis: objectives and outline

This thesis is aimed at better understanding how the human brain deals with decision-making under uncertainty. We start by asking which areas of the brain represent uncertainty in prior knowledge and current sensory information, and how they are combined (chapter 2). We continue by trying to understand the specific role of putamen and the neuromodulator dopamine in this process (chapter 3). We finish by analyzing decision-making under uncertainty in social settings, and if the way people react to it can be generalized across monetary and movement-effort situations (chapter 4).

In chapter 2, we look for neural correlates of prior and likelihood uncertainty. Indeed, although many behavioral studies have shown that people can take into account both types of uncertainty (see section 1.3), and that this is essential for optimal decision-making (see sections 1.2 and 1.3), surprisingly little is known of where in the brain these types of uncertainty are represented. We develop a decision-making task in which both prior and likelihood uncertainties are varied independently, and use fMRI to look for differential neural activations. We find neural representations of both types of uncertainty, but in very distinct locations. We discuss these results and

also refer the specific case of the putamen, which is not only more active with increased prior uncertainty, but also correlates with subjects' tendencies to sense and attend to current versus prior information

The role of the putamen in decision-making under uncertainty is further explored in chapter 3. Moreover, given that the neuromodulator dopamine is one of the main neurotransmitters from the putamen and has previously been implicated in decision-making under uncertainty (Friston, 2009, Friston et al., 2012), we also analyze the specific role of dopamine in this process. By administering the same decision-making task described above to both age-matched controls and patients with Parkinson's disease, in whom activity from the putamen is compromised due to a lack of dopamine (Kish et al., 1988, Lotharius and Brundin, 2002), we can disentangle the specific role of dopaminergic activity from the putamen. We find that some aspects of decision-making under uncertainty are conserved across patients and controls, but that patients are less sensitive to differences in likelihood uncertainty. We conclude by discussing a possible function of dopaminergic activity from the putamen in the processing of uncertainty in the current stimulus.

In chapter 4, we analyze decision-making under uncertainty in social settings. More specifically, we analyze trust, which can be considered a positive expectation in the face of social uncertainty, and ask if the way people trust can be generalized across monetary and movement-effort situations. We develop a movement effort version of the trust game, which is generally used in game theory to study trust, and compare it to the typical monetary trust game. We find no difference in subjects' decision to trust or reciprocate between both conditions. Moreover, we find a high positive correlation in subjects' behavior across conditions. We discuss these results, suggesting that the way people deal with social uncertainty may be a character trait.

Finally, in chapter 5 we summarize and discuss the findings obtained with this thesis. We also consider limitations and implications of the thesis,

and future avenues of research that could be explored in decision-making under uncertainty.

## 1.8 References

Adelson EH (1995) Checkershadow Illusion.

Anderson CH (1994) Basic elements of biological computational systems. Int J Modern Phys C 5.

Ariely D (2008) Predictably irrational: The hidden forces that shape our decisions. New York: Harper- Collins.

Bair W, Koch C (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Comput 8:1185-1202.

Barlow HB, Fitzhugh R, Kuffler SW (1957) Change of organization in the receptive fields of the cat's retina during dark adaptation. J Physiol 137:338-354.

Bastian A, Riehle A, Erlhagen W, Schoner G (1998) Prior information preshapes the population representation of movement direction in motor cortex. Neuroreport 9:315-319.

Bastian A, Schoner G, Riehle A (2003) Preshaping and continuous evolution of motor cortical representations during movement preparation. Eur J Neurosci 18:2047-2058.

Bayes T (1764) An Essay Toward Solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London 53:370-418.

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A (2008) Probabilistic population codes for Bayesian decision making. Neuron 60:1142-1152.

Beck Jeffrey M, Ma Wei J, Pitkow X, Latham Peter E, Pouget A (2012) Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. Neuron 74:30-39.

Beierholm UR, Quartz SR, Shams L (2009) Bayesian priors are encoded independently from likelihoods in human multisensory perception. Journal of Vision 9:-.

Bentham J (1780) An Introduction to the Principles of Morals and Legislation. Oxford: Clarendon Press.

Berg J, Dickhaut J, McCabe K (1995) Trust, Reciprocity, and Social History. Games and Economic Behavior 10:122-142.

Berkes P, Orban G, Lengyel M, Fiser J (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science 331:83-87.

Berniker M, Voss M, Kording K (2010) Learning Priors for Bayesian Computations in the Nervous System. Plos One 5:e12686.

Brouwer AM, Knill DC (2009) Humans use visual and remembered information about object location to plan pointing movements. Journal of Vision 9:-.

Butts DA, Goldman MS (2006) Tuning curves, neuronal variability, and sensory coding. Plos Biology 4:639-646.

Camerer CF (2003) Behavioural studies of strategic thinking in games. Trends in Cognitive Sciences 7:225-231.

Camerer CF, Fehr E (2002) Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists. SSRN eLibrary.

Carandini M, Heeger DJ (1994) Summation and division by neurons in primate visual cortex. Science 264:1333-1336.

Chami R, Fullenkamp C (2001) Trust and Efficiency. SSRN eLibrary.

Cheng K, Hasegawa T, Saleem KS, Tanaka K (1994) Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. J Neurophysiol 71:2269-2280.

Christopoulos VN, Schrater PR (2009) Grasping Objects with Environmentally Induced Position Uncertainty. Plos Comput Biol 5:-.

Cisek P, Kalaska JF (2002) Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. Journal of Neurophysiology 87:1149-1154.

Cisek P, Kalaska JF (2005) Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. Neuron 45:801-814.

Cromwell HC, Schultz W (2003) Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. Journal of Neurophysiology 89:2823-2838.

Delazer M, Zamarian L, Bonatti E, Kuchukhidze G, Koppelstatter F, Bodner T, Benke T, Trinka E (2010) Decision making under ambiguity and under risk in mesial temporal lobe epilepsy. Neuropsychologia 48:194-200.

Deneve S (2008) Bayesian spiking neurons I: inference. Neural Comput 20:91-117.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429-433.

Fehr E, Gachter S (2000) Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives 14:159-181.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.

Fiser J, Berkes P, Orban G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14:119-130.

Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431:573-578.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293-301.

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S (2012) Dopamine, Affordance and Active Inference. Plos Comput Biol 8.

Gambetta D (2000) Can We Trust Trust? In: Trust: Making and Breaking Cooperative Relations(Gambetta, D., ed): Department of Sociology, University of Oxford.

Gerardin P, Kourtzi Z, Mamassian P (2010) Prior knowledge of illumination for 3D perception in the human brain. Proc Natl Acad Sci U S A 107:16309-16314.

Glimcher PW (2003) Decisions, uncertainty, and the brain: The science of neuroeconomics. Cambridge, MA: MIT Press.

Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104-1107.

Gu Y, Angelaki DE, Deangelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201-1210.

Guseva A, Rona-Tas A (2001) Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared. American Sociological Review 66:623-646.

Helmholtz HLF (1856) Treatise on Physiological Optics: Thoemmes Continuum.

Hinde R, Groebel J (1991) Cooperation and Prosocial Behavior: Cambridge University Press.

Hinton GE, Sejnowski TJ (1983) Optimal perceptual inference. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 448-453.

Hoyer PO, Hyvärinen A (2003) Interpreting neural response variability as Monte Carlo sampling of the posterior. In: Neural Information Processing Systems vol. 15, pp 277-284, : MIT Press.

Huan Y, Rao RP (2010) Predictive coding. In: Wire's cognitive science.

Huettel SA, Song AW, McCarthy G (2005) Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci 25:3304-3311.

Kahneman D, Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. Econometrica XVLII:263-291.

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425:954-956.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227-U255.

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: Reputation and trust in a two-person economic exchange. Science 308:78-83.

Kish SJ, Shannak K, Hornykiewicz O (1988) Uneven Pattern of Dopamine Loss in the Striatum of Patients with Idiopathic Parkinson's Disease. New England Journal of Medicine 318:876-880.

Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. Quarterly Journal of Economics 112:1251-1288.

Kording K (2007) Decision theory: what "should" the nervous system do? Science 318:606-610.

Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244-247.

Kording KP, Wolpert DM (2006) Bayesian decision theory in sensorimotor control. Trends in Cognitive Sciences 10:319-326.

Krueger F, McCabe K, Moll J, Kriegeskorte N, Zahn R, Strenziok M, Heinecke A, Grafman J (2007) Neural correlates of trust. Proc Natl Acad Sci U S A 104:20084-20089.

Levitt SD, List JA (2007) What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?

Lotharius J, Brundin P (2002) Pathogenesis of Parkinson's disease: Dopamine, vesicles and alpha-synuclein. Nat Rev Neurosci 3:932-942.

Luchies LB, Wieselquist J, Rusbult CE, Kumashiro M, Eastwick PW, Coolsen MK, Finkel EJ (2013) Trust and Biased Memory of Transgressions in Romantic Relationships. J Pers Soc Psychol 104:673-694.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432-1438.

Ma WJ, Zhou X, Ross LA, Foxe JJ, Parra LC (2009) Lip-Reading Aids Word Recognition Most in Moderate Noise: A Bayesian Explanation Using High-Dimensional Feature Space. Plos One 4.

Maia TV (2009) Fear Conditioning and Social Groups: Statistics, Not Genetics. Cognitive Sci 33:1232-1251.

Maloney LT, Trommershäuser J, Landy MS (2006) Questions without words: A comparison between decision making under risk and movement planning under risk. In: Integrated Models of Cognitive Systems(Gray, W., ed) New York , NY: Oxford University Press.

McCoy AN, Platt ML (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. Nature Neuroscience 8:1220-1227.

Meyerson D, Weick KE, Kramer RM (1996) Swift trust and temporary groups. In: Trust in organizations: Frontiers of theory and research(Kramer, R. M. and Tyler, T. R., eds), pp pp. 166-195 Thousand Oaks, CA, US: Sage Publications.

Miyazaki M, Nozaki D, Nakajima Y (2005) Testing Bayesian models of human coincidence timing. Journal of Neurophysiology 94:395-399.

Morgan ML, Deangelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. Neuron 59:662-673.

Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437:1291-1298.

Osborne MJ, Rubinstein A (1994) A course in game theory: MIT Press

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223-226.

Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature 439:865-870.

Plato (360 B.C.) Republic.

Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. Neuron 51:381-390.

Riehle A, Requin J (1989) Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement. J Neurophysiol 61:534-549.

Rushworth MFS, Behrens TEJ (2008) Choice, uncertainty and value in prefrontal and cingulate cortex. Nature Neuroscience 11:389-397.

Sanfey AG (2007) Social Decision-Making: Insights from Game Theory and Neuroscience. Science 318:598-602.

Shapley R, Kaplan E, Soodak R (1981) Spatial summation and contrast sensitivity of X and Y cells in the lateral geniculate nucleus of the macaque. Nature 292:543-545.

Simpson JA (2007) Psychological foundations of trust. Curr Dir Psychol Sci 16:264-268.

Singer T, Critchley HD, Preuschoff K (2009) A common role of insula in feelings, empathy and uncertainty. Trends in Cognitive Sciences 13:334-340.

Smith MA (2001) Alhacen's Theory of Visual Perception.

Tassinari H, Hudson TE, Landy MS (2006) Combining priors and noisy visual cues in a rapid pointing task. J Neurosci 26:10154-10163.

Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. Science 307:1642-1645.

Tobler PN, O'Doherty JP, Dolan RJ, Schultz W (2007) Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. Journal of Neurophysiology 97:1621-1632.

Van Essen DC, Anderson CH, Felleman DJ (1992) Information processing in the primate visual system: an integrated systems perspective. Science 255:419-423.

Vilares I, Kording K (2011) Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Year in Cognitive Neuroscience 1224:22-39.

Wolpert DM (2007) Probabilistic models in human sensorimotor control. Human Movement Science 26:511-524.

Wu s, Amari S (2003) Neural Implementation of Bayesian Inference in Population Codes. Advances in Neural Information Processing Systems 2003:1-8.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.

# 2. DIFFERENTIAL REPRESENTATIONS OF PRIOR AND LIKELIHOOD UNCERTAINTY IN THE HUMAN BRAIN

## 2. DIFFERENTIAL REPRESENTATIONS OF PRIOR AND LIKELIHOOD UNCERTAINTY IN THE HUMAN BRAIN

**Citation:** Vilares I, Howard JD, Fernandes HL, Gottfried J and Kording KP (2012). Differential representations of prior and likelihood uncertainty in the human brain. Current Biology, 22(18): 1641-1648.

**Author Contributions:** All authors contributed in designing the study. I.V. and J.D.H. performed the experiments; I.V. analyzed the behavioral data with the help of H.F. and advice of K.K.; I.V. analyzed the imaging data with the help of J. D.H. and advice of J.G.; I.V. and K.K. wrote the paper with strong input from J. D.H., H.F. and J.G. as well.

### 2.1 Summary

**Background:** Uncertainty shapes our perception of the world and the decisions we make. Two aspects of uncertainty are commonly distinguished: uncertainty in previously acquired knowledge (prior) and uncertainty in current sensory information (likelihood). Previous studies have established that humans can take both types of uncertainty into account, often in a way predicted by Bayesian statistics. However, the neural representations underlying these parameters remain poorly understood.

**Results:** By varying prior and likelihood uncertainty in a decision-making task while performing neuroimaging in humans, we found that prior and likelihood uncertainty had quite distinct representations. While likelihood uncertainty activated brain regions along the early stages of the visuomotor pathway, representations of prior uncertainty were identified in specialized brain areas outside this pathway, including putamen, amygdala, insula, and orbitofrontal cortex. Furthermore, the magnitude of brain activity in the

putamen predicted individuals' personal tendencies to rely more on either prior or current information.

**Conclusions:** Our results suggest different pathways by which prior and likelihood uncertainty map onto the human brain, and provide a potential neural correlate for higher reliance on current or prior knowledge. Overall, these findings offer insights into the neural pathways that may allow humans to make decisions close to the optimal defined by a Bayesian statistical framework.

*“In this world nothing can be said to be certain, except death and taxes.”*

*Benjamin Franklin*

## 2.2 Introduction

Uncertainty is intrinsic to our world. For any given event, there is uncertainty in what our senses currently tell us – this is usually denoted likelihood uncertainty. There is also uncertainty in our preexisting knowledge of that event – this is known as prior uncertainty. For example, when judging the probability of rain, we combine current information obtained through our senses (are there clouds visible in the sky?) with previous acquired knowledge we possess about the chance of rain at our particular location (are we in Lisbon or London?). In such cases, the uncertainty associated with each piece of information determines how we should combine them. The combination of information gathered in the past (prior) with new information (likelihood) is critical for effective decision-making (Cox, 1946) and can thus be seen as a central objective of the nervous system.

Bayesian statistics describe how prior and likelihood information can be optimally combined as a function of their respective uncertainties to give a posterior probability estimate. The uncertainty of this optimal estimate (posterior uncertainty, or general uncertainty) is generally smaller than the uncertainty associated with either prior or likelihood alone (Glimcher, 2003). Several recent studies comparing Bayesian predictions to human behavior show that humans are close to optimal in a wide range of tasks, including estimation (Miyazaki et al., 2005), learning(Burge et al., 2008) and movement (Kording and Wolpert, 2004, Tassinari et al., 2006, Graves et al., 2010). The fact that behavior was close to the Bayesian optimal in these tasks indicates that human participants detect and use information about both prior and likelihood uncertainty. Nevertheless, in spite of a large body

of behavioral research, it is still unclear how and where these types of uncertainty are represented in the brain.

Given that uncertainty is fundamental to behavior, there is an extensive modeling literature that hypothesizes how it could be represented. However, these different theoretical models do not tend to distinguish between the representation of priors and likelihoods. Also, they differ in their predictions of where uncertainty should be represented (Ma et al., 2006, Preuschoff et al., 2006, Schultz et al., 2008, Singer et al., 2009, Fiser et al., 2010). One set of theories assumes that uncertainty is a fundamental part of the way any pair of neurons exchanges information, and thus the representation of the uncertainty of a variable is always co-localized with the representation of the variable itself (Ma et al., 2006, Fiser et al., 2010). A different set of theories assumes that there are specialized brain regions that encode and process uncertainty (Hsu et al., 2005, Preuschoff et al., 2006, Schultz et al., 2008, Singer et al., 2009). Although these theories are not necessarily mutually exclusive, they offer different predictions, and none so far has received strong neurobiological support. It thus remains unknown whether uncertainty is represented along the sensorimotor pathway or within specialized brain areas outside this pathway, and whether different forms of uncertainty have different representations.

To try to find the neural correlates of uncertainty, several studies in monkeys have analyzed how uncertainty in a stimulus can change neural firing. For instance, in a classic visual discrimination task, monkeys view a cloud of randomly moving dots and need to identify their net direction (Gold and Shadlen, 2001, Shadlen and Newsome, 2001). Varying the percentage of dots with a coherent motion demonstrated that the activity of neurons in the lateral intraparietal area (LIP) represents not only the direction of the stimulus but also the uncertainty associated with it(Shadlen and Newsome, 2001). Such studies can be interpreted as changing the likelihood. Other studies have changed the probability that even before seeing the stimulus a

monkey would have to saccade to a given target, and found that this modulated neuronal activity (Basso and Wurtz, 1997). These studies can be interpreted as changing the prior (Gold and Shadlen, 2001). However, all of these studies are based on relatively simple oculomotor tasks, with a focus on very specific brain areas. It remains poorly understood whether neural representations of uncertainty are also encoded elsewhere in the brain, and whether the findings in monkeys translate to other experimental tasks and settings in humans.

Interesting recent human studies in neuroeconomics have started to examine how uncertainty about reward is represented. These studies show that increased uncertainty about whether a reward is going to be obtained (risk) correlates with increased activations in the orbitofrontal cortex (OFC)(Critchley et al., 2001, Tobler et al., 2007) , midbrain (Preuschoff et al., 2006), cingulate cortex (Critchley et al., 2001, Behrens et al., 2007) , and insula (Huettel et al., 2005, for a review, see Rangel et al., 2008, Schultz et al., 2008). However, in all these cases uncertainty was treated as one single parameter, and thus general uncertainty was perfectly correlated with prior or likelihood uncertainty, which were not distinguished as separate sources of uncertainty. Thus, although in humans there is an increasing knowledge of where uncertainty in reward is represented, it is still unclear where prior and likelihood uncertainty are, and if these areas coincide with the areas involved in uncertainty in reward.

Here we devised a sensory-motor decision-making task in which human subjects could use both prior and likelihood to estimate positions of hidden visual targets. Uncertainty was systematically varied on each trial in a two-by-two factorial design, such that two of the conditions were matched for performance accuracy but differed in the relative balance of prior and likelihood uncertainty. By combining event-related functional magnetic resonance imaging (fMRI) approaches with computational models of behavioral performance, we were able to characterize the neural

representations of the two kinds of uncertainty while controlling for confounds related to expected reward. The behavioral results indicate that subjects are adept at using both kinds of uncertainty to optimize performance, in keeping with Bayesian predictions. The imaging results suggest that likelihood uncertainty is primarily represented in the early stages of the sensorimotor network, while prior uncertainty is represented in limbic and paralimbic decision-related areas outside of traditional sensorimotor pathways. Together these findings suggest fundamentally different representations by which prior and likelihood uncertainty in a decision-making task map onto the human brain.

## 2.3 Materials and methods

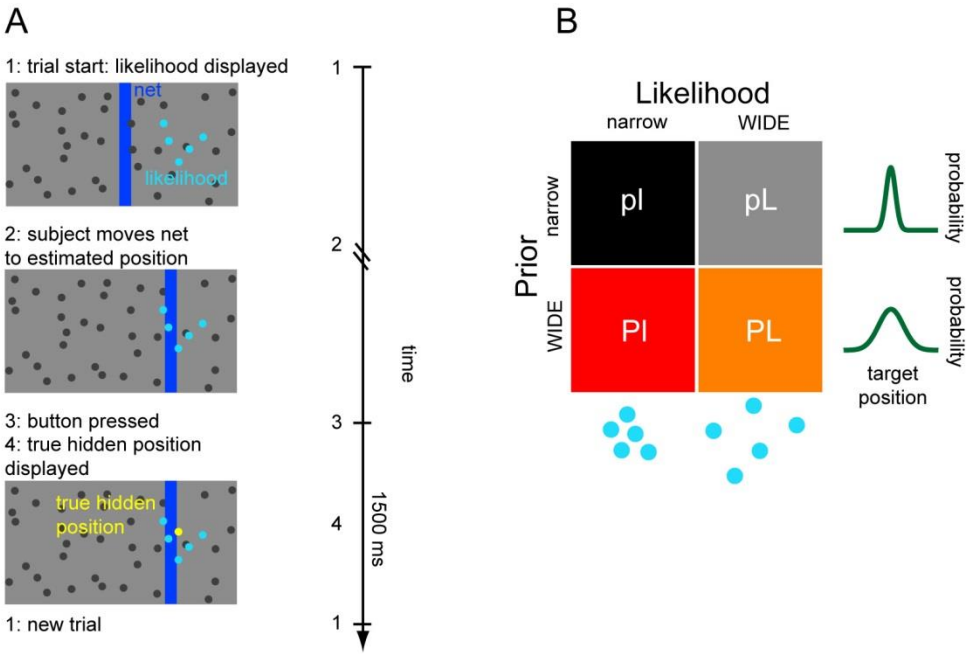


**Figure 2.1 - Experimental setup.** (A) Illustration of the task. Subjects guess the position of a hidden target (the "coin", represented by the yellow dot) using a net (vertical blue bar) which they can displace horizontally. At the onset of each trial, subjects receive noisy information about the position of the hidden target in the form of a set of 5 blue dots (the likelihood). Subjects then move the net to the guessed position and press the mouse button to confirm their choice, after which the true target position is displayed. A new trial then begins 1500 ms later. Left: illustration of the computer display that was presented to the subjects. Right: typical time course of a trial. (B) The 4 conditions of the experiment. The experiment consisted of a two-by-two factorial design, with two types of prior (p=narrow prior; P=wide prior) and two types of likelihood (l=narrow likelihood; L=wide likelihood). The wider conditions are the ones with more associated uncertainty.

**Subjects.** Twenty-seven healthy subjects (12 women; age range 19-35 years; mean age=27 years) participated in the experiment. Of these, 17

participated in the fMRI experiment, from which data from 15 were used (9 women). All participants were right-handed, had normal or corrected-to-normal vision, were naive to the goals of the experiment, signed consent forms and were paid to participate. Subjects that only performed the behavioral part of the experiment were paid $20. Subjects that performed both the behavioral and the fMRI parts of the experiment were paid $70. All protocols were approved by the Northwestern University IRB.

**Behavioral task.** Subjects performed a decision-making task, which consisted of guessing the position of a hidden coin on a screen, in a task similar to one described in prior studies (Tassinari et al., 2006, Berniker et al., 2010). Subjects were told the cover story of a coin being tossed into a pond and informed that their task was to guess where the coin had fallen. They could not see the coin, but they could see 5 blue dots that were the "splashes" produced by the coin falling in. They were told that the person who threw the coin aimed, albeit imperfectly, at the center of the screen (mean of prior). They were also told that, between blocks, the thrower changed, and the new one might be better or worse at throwing (i.e. they were indirectly informed that the variance of the prior changed). To estimate the coin position, subjects could use (although they were never explicitly told so) both the coin position's likelihood, obtained from the "splashes", and its prior (the distribution of previous coin locations). There was no temporal deadline.

**Stimuli.** The position of the coin was drawn from a Gaussian distribution, centered on the center of the screen with a standard deviation (std) that was either low ($\sigma_p$=2.5% of screen width) or high ($\sigma_P$=8.5% of screen width). This distribution was the *prior* of the experiment. Subjects were given the mean

of the prior ("the coin throw is aimed at the screen center") but not its variance, which they could only estimate from the distribution of previous coin throws. The standard deviation of the prior was kept constant within blocks, but changed across blocks. On every trial, a cluster of five dots was shown on the screen. The x-position of each of these dots was drawn independently from a second Gaussian distribution in which the mean was the coin's horizontal location on that trial and the standard deviation was either low ($\sigma_l$=6%) or high ($\sigma_L$=15%). The distribution of these five dots defined the *likelihood*. The std of the likelihood was varied pseudo-randomly from trial to trial but counterbalanced across trials. We made the std of the likelihood vary pseudorandom from trial to trial so that subjects could not predict *a priori* the overall uncertainty that the trial would have. In total there were thus four conditions: low prior uncertainty and low likelihood uncertainty (pl); low prior uncertainty and high likelihood uncertainty (pL); high prior uncertainty and low likelihood uncertainty (Pl) and high prior uncertainty and high likelihood uncertainty (PL).

The likelihood dots were displayed as blue dots against a gray background (see Figure 2.1-A). The coin appeared as a yellow dot. The background was light grey with darker grey dots dispersed randomly. These darker grey dots, however, never changed location. This particular background was chosen to minimize possible effects on the fMRI data related to edge effects. All dots had the same diameter. The screen units were normalized between 0 (the left edge) and 1 (the right edge). Stimulus presentation, for both the lab portion and the scanning session, was performed using Matlab 7.4.0 (MathWorks, Natick, MA).

**Choosing the parameters – matching of performance.** The std values were chosen such that, on average, people would perform with identical accuracy in the pL and the Pl conditions. For that, different std values were

tested on an individual level for a small number of pilot subjects, until these subjects performed with identical accuracy in the pL and Pl conditions. This gave us a specific set of parameters which we then used for the main experiment. We decided to use the same set of parameters for all following subjects because: (1) to make subject-by-subject matches of performance by finding the appropriate parameter values for them would require potentially more than triple the experimental time, and even then there would be significant remaining uncertainty. In order to maintain a tolerable experimental duration that is not overly taxing for subjects, matching on the basis of group averages is a distinct advantage. (2) Having different parameter values for different subjects would statistically complicate group comparisons.

**Procedure: Behavioral session.** Experimental blocks: At the onset of each trial, five blue likelihood dots were shown on the screen, where they remained until the end of the trial. Subjects had to move a blue vertical bar (the "net") with a trackball mouse to estimate the coin position and press the right mouse button once they made their decision. Subjects could take as long as they wanted to decide where to place the net but had to wait for at least 1.26 seconds. After they pressed the button, the true position of the coin was revealed and subjects would get one point added to the score if the coin was inside the net. The cumulative score across the experiment was shown to the subjects at the end of each trial. A new trial would then begin 1.5 seconds later (see Figure 2.1-A). Given that the net covered the entire height of the screen only the horizontal location was relevant, making this a one-dimensional estimation task. Subjects completed four blocks of 150 trials each for a total of 600 trials per experiment. Each block was characterized by a particular std of the prior. Half of the subjects started with the low prior std condition and the other half started with the high prior std condition. After the four blocks, subjects also performed 100 trials of a

control block. Each subject thus had to perform 700 trials. Each trial took an average 4 seconds and the total experiment lasted on average approximately one hour.

Control Block: The control block was performed immediately following completion of the four experimental blocks, on those subjects that were rapid to complete the main block (n=11). The task was identical to the main experiment, with the only difference being that the coin's location was shown at the onset of each trial and could be seen throughout, so there was no uncertainty about its position. The 100 trials comprising the control block were selected by randomly sampling 25 trials per condition from the main experiment blocks. Each trial of the control block repeated one of those sampled experimental trials, showing the same likelihood dot display that was shown in the experimental trial and using as the coin position the actual position to which the person moved the net at that trial. As in the main experiment, subjects were awarded one point for successfully moving the net to the coin's position.

**Procedure: fMRI session.** From the 27 subjects recruited for the behavioral task, 17 were asked to return and perform the same task while undergoing fMRI. The choice of which subjects would also do the fMRI part of the experiment depended mainly on subjects and scanner times availability. If, however, more than one subject was available for the same scanner time slot, then the one chosen to participate would be the one in which more behavioral difference between the different conditions was observed. This was done in order to maximize statistical power. Note that for the fMRI part of the experiment we were interested in the neural representations of prior and likelihood uncertainty and not if they were or not behaving in a Bayes-like way. The task was identical to the one performed in the behavioral experiment, minimizing the need for new learning. Because it was the

second time they performed the same task, it was assumed that after the start of each block they rapidly converged to the correct prior(Berniker et al., 2010). The four experimental blocks were performed in succession with a 2-minute break in between each one. At the end of the experiment a T1-weighted anatomical scan was acquired. Data from two of the 17 subjects could not be included in the analysis due to technical problems in the scanner, and for one other subject, data from only two of the four blocks were acquired due to technical problems (but because the 2 blocks had different prior uncertainties, the four types of trials were equally represented). Due to technical issues and time constraints, data from the control block were only obtained in 11 of the subjects. Thus, in total, data from 15 subjects were used in the main experiment and data from 11 of them were used in the control task.

**Data Analysis:**

**Bayesian modeling of behavior.** In order to perform the task successfully, for every trial the subjects should place the net in the most likely location of the hidden coin. Bayes rule provides an optimal way to estimate this location(Kording and Wolpert, 2004):

$$X_{est}=\sigma^2_L/(\sigma^2_L+\sigma^2_P)\mu_P+\sigma^2_P/(\sigma^2_L+\sigma^2_P)\mu_L \qquad (1)$$

where $X_{est}$ is the estimated position of the coin, $\sigma^2_L$ and $\sigma^2_P$ are the variances of the likelihood and of the prior, respectively, and $\mu_L$ and $\mu_P$ are the respective means. For our experiment, the real $\mu_P$ (the mean of the prior) is always a constant, in this case 0.5 (the center of the screen). The mean of the likelihood ($\mu_L$) for each particular trial can be considered the centroid of the cloud of dots, and it changes from trial to trial. We can then make a linear regression of the subject's estimated coin position, $X_{est}$, as a function of the centroid of the cloud of dots (see Figure 2.2-A). The slope of

this linear regression characterizes how much the subject is weighting the likelihood information, and, if people perform according to the optimum prescribed by Bayesian statistics, its value should be equal to the perceived $\sigma^2_P/(\sigma^2_L+ \sigma^2_P)$. A slope of zero suggests that subjects do not take into account likelihood information, and a slope of one suggests that subjects only use likelihood information. A slope between zero and one indicates that subjects are using information from both prior and likelihood, and the larger the slope the more they rely on the likelihood and less on the prior.

*Subjects' mean of the prior*: The intercept from this linear regression ($\beta_0$) should be equal to the perceived $\sigma^2_L/(\sigma^2_L+ \sigma^2_P)\mu_P$. Thus, to estimate the mean of the prior that the subjects were using, and given that $\sigma^2_L/(\sigma^2_L+ \sigma^2_P)=1-\sigma^2_P/(\sigma^2_L+ \sigma^2_P)=1-$ *slope*, subjects mean of the prior can be calculated by: $\mu_P= \beta_0/(1-$ *slope*$)$.

*Subjects' average reliance on likelihood vs. prior information:* For the correlation between fMRI signal and individual behavior tendencies, the slopes associated with each trial type were averaged to get an estimate of an individual's average reliance on likelihood information.

*Subjects' performance:* The proportion of correct trials, for each condition, was calculated as the number of trials in which the subject accurately guessed the position of the coin, divided by the total number of trials in that condition.

For the behavioral calculations, the first 50 trials of every block were discarded to minimize the effect of learning.

*Optimal values:* The Bayesian optimal values for the slopes can be obtained from $\sigma^2_P/(\sigma^2_L+ \sigma^2_P)$, where $\sigma^2_P$ is the variance associated with the prior (with $\sigma^2_p=0.025^2$ or $\sigma^2_P = 0.085^2$ in unit-less screen coordinates) and $\sigma^2_L$ is the variance associated with the likelihood, which can be estimated by $\sigma^2_L$= variance (cloud of dots)/ number of dots (so in our case, $\sigma^2_l=0.06^2/5$ or $\sigma^2_L = 0.15^2/5$).

*Instantaneous Bayesian slope:* We can also take Equation 1 and rearrange it to give us the "instantaneous Bayesian slope", i.e. how much the subject is taking the likelihood information into account on each trial. If we denote this slope by *s*, then we can rewrite the equation as:

$$X_{est}=(1 - s )\mu_P+ s \mu_L \qquad (2)$$

Which gives us the slope per trial :

$$s = (X_{est} -\mu_P )/(\mu_L -\mu_P ) \qquad (3)$$

This slope will also give us 0 if $X_{est} =\mu_P$ (i.e. if the subject puts the cursor at the mean of the prior) and 1 if $X_{est} =\mu_L$ (i.e. if the subject puts the cursor at the mean of the likelihood). However, this slope value gets unstable when $\mu_L$ and $\mu_P$ are very close together (i.e. when $\mu_L$ -$\mu_P$ approaches zero). Therefore, in the "instantaneous Bayesian slope" model we used instead the arctangent of this measure, namely $arctg\,(-0.5 + s_L)\,/\pi + 0.5$, so to have a measure bounded by 0 and 1.

*Sensitivity to prior change*: This is measured as the average absolute difference between the slope at the end of one block and the slope at the beginning of the new block (which has a different prior uncertainty). To

calculate each slope the last/first 20 trials of the block were used. To capture real sensitivity to prior change, and not learned effects, and also to minimize any circularity, the values used were the ones obtained during the lab portion of the experiment (when they were first exposed to the task).

**fMRI data acquisition and processing.** Whole-brain imaging was performed on a 3-Tesla Siemens Trio MR scanner, using a 32-channel head coil. Functional data were acquired using a gradient-echo echo-planar scanning sequence (repetition time = 2.76 sec, echo time = 20 ms; 44 axial slices; matrix size = 128 × 120 voxels; field-of-view = 220 × 206 mm; in-plane resolution = 1.72 × 1.72 mm; slice thickness = 2 mm; gap = 1 mm; flip angle = 75°), using a 32-channel head coil. All functional images were acquired in a single session that lasted about one hour with some variation due to response times. Participants were placed in a light head restraint within the scanner to limit head movement. Visual stimuli were projected on a screen and were viewed through a mirror attached to the head coil. After the acquisition of functional images, a high-resolution T1-weighted anatomical scan was obtained, which were later coregistered to the mean functional image of each subject, normalized to a standard template image, and averaged across subjects to aid in localization of observed brain activations. All of the fMRI images shown throughout the paper have been overlaid on this averaged T1 (in which only the T1 scans of 8 of the subjects were used, due to technical difficulties).

The first six brain images acquired in each fMRI run were discarded to allow for T1-weighted signal equilibration. Scan onset times varied relative to stimulus onset times (trial start). Image processing and statistical analyses were performed using Statistical Parametric Mapping software (http://www.fil.ion.ucl.ac.uk/spm). All functional images for a given subject were realigned to the first volume of the first run to correct for head

movement, normalized to a standard echo planar imaging (EPI) template, and spatially smoothed using a 6-mm (full width at half-maximum) Gaussian kernel.

**fMRI data analysis.** For every participant, a standard rapid event-related fMRI approach was used, in which the onset of each trial type (4 different trial types in our experiment) was convolved with a canonical hemodynamic response function and then regressed against the measured fMRI signal (Josephs and Henson, 1999). Because each trial type appeared in a random order and had a different duration (controlled by the subject, but not less than 2.76s), the fMRI responses to the rapidly presented events could be separated without the need to wait for the hemodynamic response to return to baseline(Josephs and Henson, 1999) after each trial. The onset times were generally considered as the beginning of each trial, except for the reward/no-reward model, in which the onset times corresponded to the time in which the subject had already made a response and was shown the real position of the coin. The general linear model was first estimated at the single-subject level and it contained: one regressor for each of the four trial types (pl, pL, Pl, PL), separated by prior-type block (therefore, in each block, only two trial types existed); six nuisance regressors per block that corresponded to participant-specific head-movement parameters; and regressors encoding the average BOLD response at each of the experimental blocks. For the parametric likelihood uncertainty model, the "instantaneous Bayesian slope" model and the reward model all blocks were concatenated as one combined block, and 3 additional nuisance variables were included to minimize block/session effects. The parameter estimates (beta values) that were estimated from this general linear model reflect the strength of covariance between the canonical hemodynamic response function and the brain activation for a given condition(Friston et al., 1995).

Contrast images, derived from a pair-wise contrast between each trial type and an implicit baseline, were then taken to a second-level group analysis into a repeated-measures ANOVA ("flexible factorial"; to study the effects of each trial type) or a paired t-test (to study potential reward effects) or a simple t-test (to study potential parametric likelihood uncertainty effects or the potential existence of an area that tracked the "instantaneous Bayesian slope"). Throughout, the statistical threshold was set to p=0.05 family-wise error corrected for multiple comparisons over the entire brain volume. All areas that are reported (see also Sup. Tables in the Appendix of this thesis) survive the corrected thresholds at the peak and/or cluster level. The anatomical localization of the local maxima was aided with reference to a human brain atlas (Duvernoy, 1999).

*Cross-validation*: Although the voxel selection procedure outlined above does not give rise to any circularity, if we were to simply plot the values for percent signal change using the model data that generated our brain maps there would be some circularity between voxel identification from voxel plotting(Jazayeri and Shadlen, 2010). In order to minimize such circularity, the values for percent signal change (PSC) shown in Figure 2.3-B and Figure 2.5-B were extracted using a cross-validation leave-one-out procedure (similar to what is reported in Ref. Glascher et al., 2010): the flexible-factorial from the second-level analysis was re-estimated but leaving out one subject. For the different regions of interest, the new maximum coordinates for this second-level analysis were extracted. The beta values for these new voxels were extracted from the left-out subject, thereby minimizing subject independence between voxel selection and voxel plotting. This procedure was iterated fifteen times, each time with a different subject "left-out". The statistical threshold was set at $p<0.05$ for the cross-validation procedure.

We also performed a two-fold cross-validation method, in which instead of taking away one subject at a time, we randomly divided the subject pool in

half, used half for localization and extracted the beta values from the remaining half of the subjects. The same basic results were found than in the leave-one out cross-validation procedure: there was a significant effect of prior and not of likelihood in the right and left putamen, right amygdala, left insula and right OFC; and a significant effect of likelihood and not prior in the occipital cortex.

Note that, because there was no true zero baseline (e.g. a condition with no prior uncertainty), what matters is the relative difference in the beta values/percent signal change between the different conditions.

## 2.4 Results

In this study we wanted to know where prior and likelihood uncertainty are represented in the human brain. To this end, we developed a visual decision-making task in which subjects had to guess the position of a hidden target (a “coin”) on a computer screen (see Figure 2.1-A and (Tassinari et al., 2006, Berniker et al., 2010)). Subjects were given noisy visual information about the target position in the form of a dot-cloud drawn from a Gaussian distribution centered at the true target position. To successfully estimate the position of the target, subjects could use both the likelihood, obtained from the displayed dots, and the prior, obtained from the distribution of previous target positions. Uncertainty in the likelihood varies with the dispersion of the displayed dots. This dispersion varied randomly from trial to trial, and thus could not be predicted beforehand. The average position of the target (the mean of the prior) was the middle of the screen, and subjects could estimate its uncertainty (prior uncertainty) from the distribution of target positions in previous trials (see *2.3 Materials and methods* for details). Given that subjects had ample experience with the task from the behavioral experiment, they quickly acquired the prior. Successful estimates of the position were rewarded with points, which had motivational significance to the subjects (see *2.4.2 Additional results* and Figure 2.8). The conditions comprised a two-by-two factorial design (Figure 2.1-B), with two levels of prior uncertainty (wide, more uncertain prior: **P**, and narrow, less uncertain prior: **p**) and two levels of likelihood uncertainty (wide likelihood: **L**, and narrow likelihood: **l**). Varying the uncertainty in the prior and in the likelihood independently, together with fMRI imaging, allowed us to find where they are represented in the brain.

### 2.4.1 Main results

#### Prior and likelihood uncertainty affect behavior

We first wanted to know if variations of prior and likelihood uncertainty in our task influence the estimation behavior. If subjects ignore the prior information and rely only on the current sensory feedback (i.e. likelihood information) then the weight of sensory feedback (slopes of Figure 2.2-A) in the estimation should be one. On the other hand, if subjects rely only on their prior knowledge and ignore likelihood information, then the weight of the sensory feedback should be zero. Use of this metric demonstrated that scanned subjects relied on both prior and likelihood information (0<slope<1; Figure 2.2-B). Data obtained from all subjects (both those who were scanned and those who were not) showed the same effects (Figure 2.9-A). Furthermore, subjects relied more on the likelihood information as the prior uncertainty increased and as the likelihood uncertainty decreased (main effect of prior, $F_{1,43}=207$, $p<10^{-6}$; main effect of likelihood, $F_{1,43}=35$, $p<10^{-6}$, ANOVA repeated measures (r.m.)). Qualitatively, such behavior would be expected if they used a Bayesian strategy (Vilares and Kording, 2011). Thus, in our experiment, subjects utilized knowledge of both prior and likelihood uncertainty for perceptual decision-making.

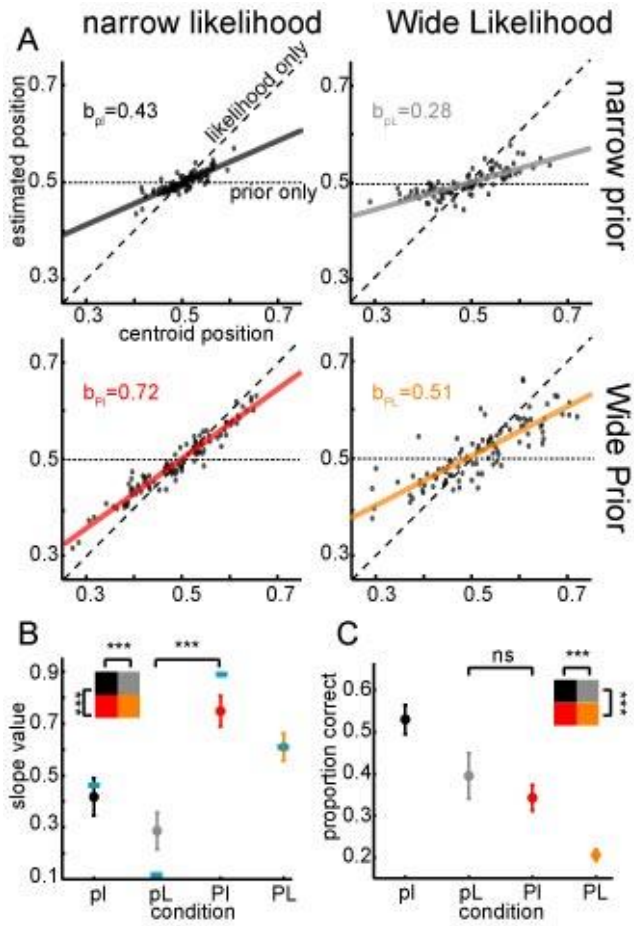


**Figure 2.2. Behavioral results.** (A) Estimates of the target position for one representative subject are shown as a function of the centroid of the displayed dots (likelihood). Displayed next to the graphs is the slope value of the linear regressions (solid line). The dashed line represents what the linear regression would look like if the subject only used likelihood information (slope=1). The horizontal dotted line represents a potential situation in which only prior information is taken into account (slope=0, localized at 0.5 which is the middle of the screen and the mean of the prior). (B) Average slope of the linear regression for the behavior of the 15 subjects during the scanning session, separated by condition. The slope quantifies the degree to which subjects rely on the current visual stimulus (likelihood) vs. the prior. The small blue rectangles represent the optimal Bayesian values (see *2.3 Materials and methods* for details). (C) Average proportion of trials in which the subject accurately guessed the position of the target, separated by condition. Error bars in

*(B)* and *(C)* represent 95% confidence intervals for the mean (s.e.m.). Inset shows that there is a significant effect of both prior and likelihood uncertainty. *n.s.*= non-significant, p>0.05; *** significant, p<0.001. See also Figure 2.9.

Both kinds of uncertainty may be expected to change the precision of subjects' estimates and thus their expected task performance. Not surprisingly, performance was better when each of the uncertainties was lower (Figure 2.2-C, main effect of prior, $F_{1,43}$=161, $p<10^{-6}$; main effect of likelihood, $F_{1,43}$=84, $p<10^{-6}$, ANOVA r.m.). Importantly, despite the fact that the slopes, and thus the relative weighting of the uncertain sources of information, differed significantly between the pL and Pl conditions (Figure 2.2-B, $p<1x10^{-4}$, W=0, Wilcoxon signed rank test), estimation performance between these two conditions was matched (Figure 2.2-C and Figure 2.9-C, p>0.05, W=31, Wilcoxon signed rank test). Even testing estimation performance on each individual subject did not reveal any significant difference between these two conditions (p>0.05 for each subject, corrected for multiple comparisons, comparing two proportions test). Therefore, as noted below, a comparison of these two conditions should provide a robust way to infer imaging-based differences between prior and likelihood uncertainty, while minimizing potential confounds due to differences in general uncertainty, performance, or reward expectation.

**Prior and likelihood uncertainty have distinct neural representations**

The uncertainty associated with the prior affects behavior (Figure 2.2-B). We thus wanted to ask if and where it affects BOLD activities (see *2.4.2 Additional results* for details). We found that areas more active with increased uncertainty in the prior include putamen, amygdala, insula, and OFC ($p<0.05$ whole-brain corrected; family-wise-error (FWE) level; Figure 2.3-A). Cross-validated condition-specific activation profiles (Figure 2.3-B) demonstrate that the activation in each of these regions is specifically related with prior uncertainty, and not likelihood uncertainty. Please note that, because in our design there is no "zero prior uncertainty" condition no real baseline exists, and thus what is relevant in the activation profiles is the relative difference between conditions. There was a significant main effect of prior uncertainty (all regions depicted are significant at $p<0.01$, $F>9$; except left amygdala, with a $p=0.059$, $F_{1,43}=3.76$; ANOVA r.m.), with no significant main effects of likelihood ($F_{1,43}<0.73$, $p>0.05$; ANOVA r.m.). Even if we use a two-fold cross-validation, the same results still hold (see *2.4.2 Additional results* for details). We did not find any imaging evidence of an interaction between priors and likelihoods (for additional tests and controls, see *2.4.2 Additional results* and Figure 2.11-A). Areas more active with low prior uncertainty (i.e. more active with increased precision/confidence) encompassed the caudate nucleus, prefrontal cortex and areas adjacent to the anterior cingulate cortex (see Figure 2.4). Together, these results suggest that wide regions of the brain, primarily outside of the traditional sensory-motor pathway, encode prior uncertainty.

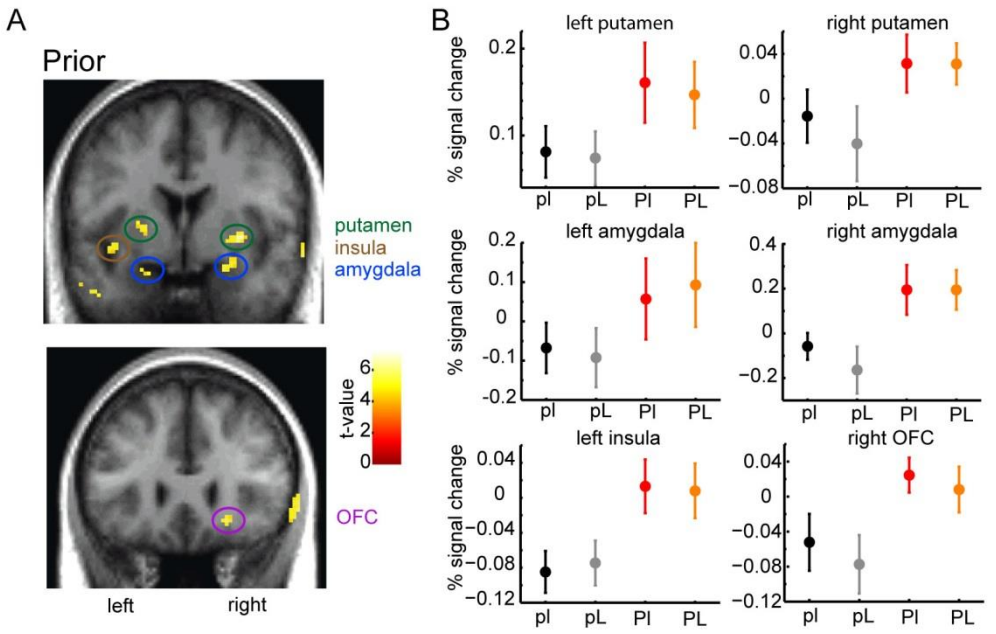


**Figure 2.3. Brain regions more active during wide (more uncertain) prior conditions.** (A) Stronger activations associated with high (vs. low) prior uncertainty were seen bilaterally in the putamen, amygdala, insula (top; y=0), and OFC (bottom; y=26). Functional activations are overlaid on coronal sections of the average of each subject's T1-weighted structural brain scan (display threshold at p<0.0001 unc., minimum 10 voxels; n=15). Activity in the right insula and left OFC appears at a less stringent p-value (p<0.001; not shown). In this and all subsequent figures, the right side of the brain corresponds to the right side of the image. (B) Percent signal change (PSC) by condition in the areas represented in *(A)*. Data were extracted from the peak (most significant) voxel in each cluster using a leave-one-subject-out cross-validation procedure (see *2.3 Materials and methods* for details). Plots represent the subject-averaged parameter estimates converted to PSC for the four conditions (n=15). Error bars represent standard error of the mean (*s.e.m.*).

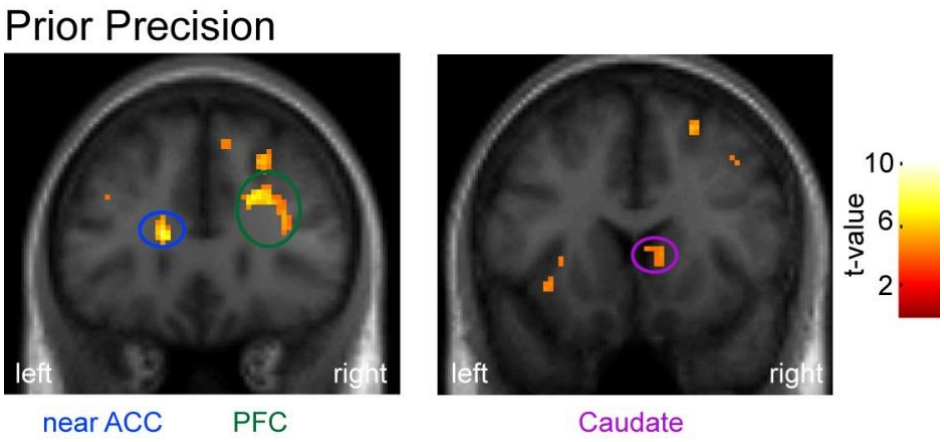


**Figure 2.4 - Brain regions more active during narrow (less uncertain) prior conditions.** Stronger activations associated with low (vs. high) prior uncertainty were seen near the anterior cingulate cortex, the prefrontal cortex (left panel; y=29) and the caudate (right panel; y=14). Functional activations are overlaid on coronal sections of the average of each subject's T1-weighted structural brain scan (display threshold at p<0.0001 unc., minimum 10 voxels; n=15). The right side of the brain corresponds to the right side of the image.

Given that activity related to prior uncertainty encompassed a wide set of regions, we wanted to know if the uncertainty associated with the currently presented stimulus (the likelihood uncertainty) leads to activity in the same areas. We should expect that a neural representation exists, given that likelihood uncertainty also affected behavior (Figure 2.2-B). Areas more active with high vs. low likelihood uncertainty were localized to bilateral regions of superior occipital visual cortex (p<0.05 FWE corrected; see Figure 2.5) and nowhere else in the brain at this threshold (see *2.4.2 Additional results* and **Figure 2.11**-B for controls). Taking advantage of the fact that the degree of likelihood uncertainty differed on every trial (where greater dot dispersion corresponded to higher visual variance and higher likelihood uncertainty) and that the between-trial variability in dot position produces some overlap between the low and high likelihood uncertainty

groups (see *2.4.2 Additional results* and Figure 2.10), we also implemented a parametric fMRI model to test whether dot-dispersion variance modulated the same areas of occipital cortex on a trial-by-trial basis. This model revealed the same bilateral activations in the superior occipital visual cortex. These activations increased parametrically with the increase in likelihood uncertainty, had a higher level of significance and extended all the way to the calcarine sulcus and lingual gyrus ($p<0.05$ FWE corrected; Figure 2.5-C). Likelihood uncertainty in our task thus seems to affect BOLD signal mainly in areas corresponding to the early stages of the visuomotor pathway.

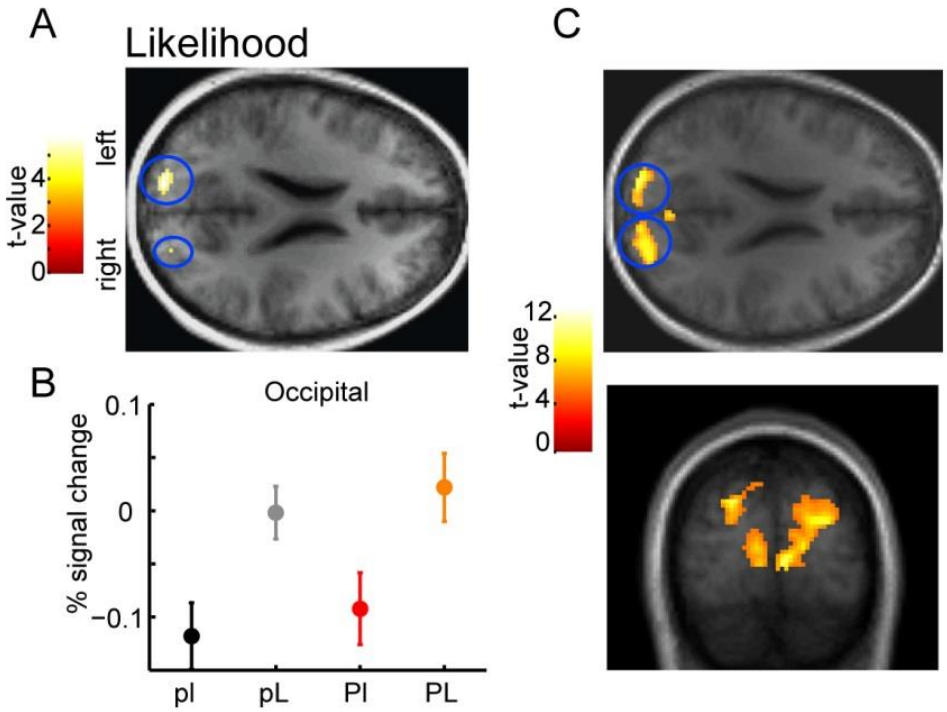


**Figure 2.5. Brain regions more active during higher likelihood uncertainty conditions.** (A) Stronger activations associated with high (vs. low) likelihood uncertainty were seen bilaterally in the superior occipital cortex (z=22). Functional activations are overlaid on the axial section of the subject-averaged scan (n=15). (B) Condition-specific percent signal change in the left occipital area shows a main effect of likelihood uncertainty significant at $p<0.0001$, with no significant main effect of prior ($p>0.05$). Data were extracted using a leave-one-subject-out cross-validation procedure. The same results hold if using a two-fold cross-validation procedure. The data plot represents the average parameter estimates (±s.e.m., cross-validated) converted to percentage signal change for the four conditions. (C) Brain regions parametrically correlated with higher likelihood uncertainty. Stronger activations associated with a parametric increase in likelihood uncertainty (standard deviation of the displayed dots at each trial) were seen bilaterally in the superior occipital cortex, extending down all the way to the calcarine sulcus and lingual gyrus (y=-88; z=22). Functional activations are overlaid on the axial (up) and coronal (down) sections of the subject-averaged scan (for all functional activations shown, display threshold is at $p<0.0001$ unc., minimum 10 voxels; n=15).

**Correlations across subjects between prior-related behavior and activations**

Thus far, we have analyzed how the two factors, prior and likelihood uncertainty, affect activities in the brain, and identified a small number of key regions. For those analyses we did not use observed behavior but rather the experimentally and mathematically defined uncertainties. However, if any of these brain regions are involved in the decision-making process then we should expect them to be correlated with behavioral variance across subjects (see also *2.3 Materials and methods* and *2.4.2 Additional results*).

We hypothesized that subjects who were able to better detect the changes in prior uncertainty (i.e. those who changed more their behavior when the prior uncertainty changed) could also show higher differences in brain activity between the two prior conditions. This is indeed what was observed for both the right and left putamen (Figure 2.6-A:B, $r=0.65$ and $r=0.61$, $p<0.02$, Spearman correlation) and the OFC (Figure 2.6-C, $r=0.66$, $p<0.01$, Spearman correlation). The other reported areas (namely, amygdala and insula) did not show a significant correlation. These results suggest that, besides tracking prior uncertainty, putamen and OFC might be directly related with behavioral change.

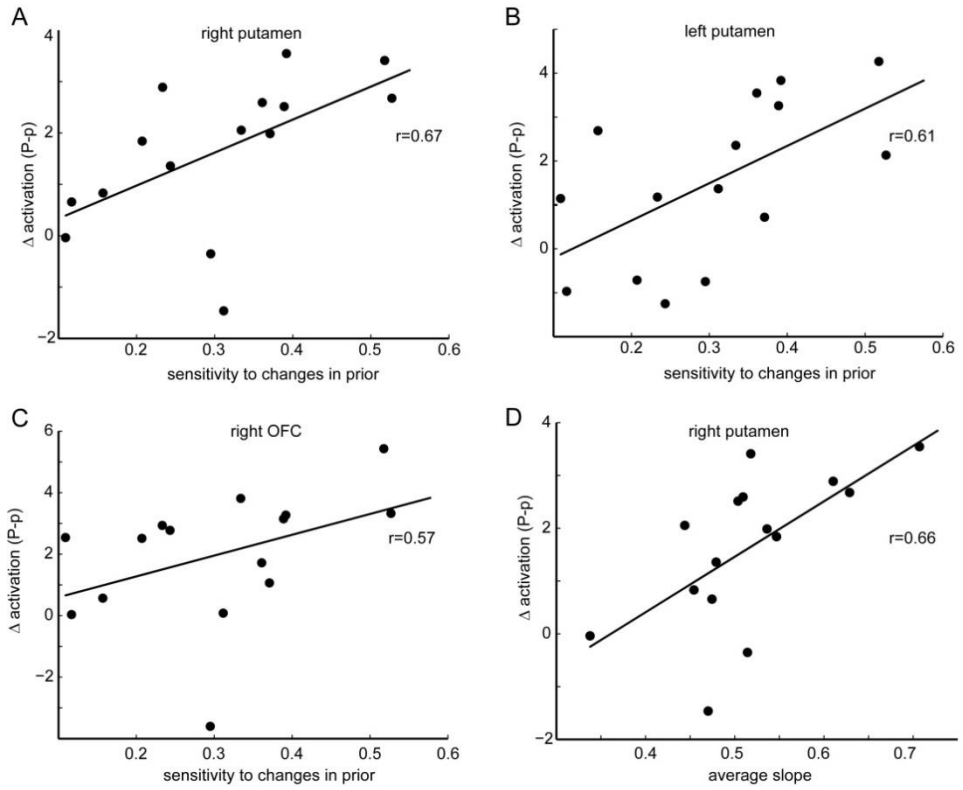


**Figure 2.6. Relationship between prior uncertainty-related brain activity and behavioral measures.** (A-C) Prior uncertainty differences in activation for individual participants at the right putamen (A; peak at 30, -2, -2), left putamen (B; peak at -22,0,4) and OFC (C; peak at 22,26,-12) are regressed against individuals' sensitivity to prior change. Presented on the graphics are the respective Spearman correlations. (D) Prior uncertainty differences in activation for individual participants at the right putamen (peak at 30, -2, -2) are regressed against the average weight given to likelihood vs prior information (the slope). Presented on the graphic is the respective Spearman correlation.

We also wanted to look for a behavioral measure that could represent a specific "personality trait". The average slope of a subject (averaged over the four conditions) can provide such a measure, as a higher value indicates that subjects on average tend to rely more on current

sensory information and less on prior knowledge, and vice versa. Indeed, as we have seen above, subjects varied behaviorally in how much, on average, they relied on prior information (see Figure 2.2), and also varied in their mean fMRI activation response to uncertainty in prior-related brain areas (see Figure 2.3). We found that this average slope was positively correlated with the degree of differential brain activation in the right putamen (Figure 2.6-D, $r=0.66$, $p<0.01$, Spearman correlation). Note that the effect in the putamen (in terms of difference between conditions) is actually the highest compared to the other brain areas (see Sup. Table S1). This finding raises the possibility that a general higher involvement of the putamen in the task, potentially by signaling greater prior uncertainty, may enhance learning from new sensory information, and therefore subjects in whom the putamen is more engaged tend to rely more on new than in prior information.

**Online representation of Bayesian slopes**

We wanted to ask how the brain computes prior versus likelihood information online in order to arrive at a sensorimotor decision. For that decision, at each trial the subject needs to decide how much to weigh the likelihood relative to the prior. If subjects behave as predicted by Bayesian statistics, then this slope can be estimated for each trial (see *2.3 Materials and methods* for details). Testing for a correlation between this "instantaneous Bayesian slope" and BOLD activation, we found significant decision related activations in the prefrontal cortex (PFC), roughly in Brodmann area 9 ($p<0.0001$ unc., see Figure 2.7). This finding suggests a possible role of prefrontal cortex in combining prior and likelihood information to estimate the target's position.

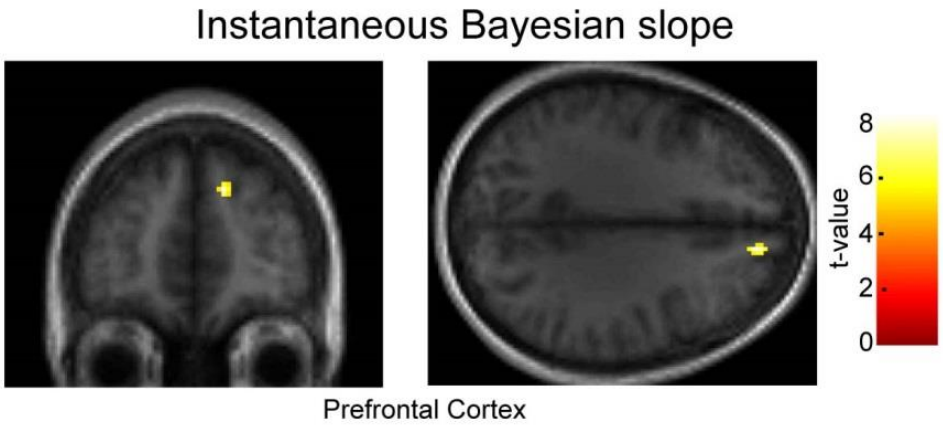


**Figure 2.7 - Brain regions parametrically correlated with the instantaneous Bayesian slope.** Stronger activations associated with a parametric increase in the instantaneous Bayesian slope (a measure that indicates how much the subject weights prior and likelihood information, at each trial) were seen in the right superior medial prefrontal cortex (y=48; z=32). Functional activations are overlaid on the coronal section (left) and axial (right) sections of the subject-averaged scan (display threshold at $p<0.0001$ uncorrected, minimum 10 voxels).

### 2.4.2 Additional results

#### Comparison between prior and likelihood uncertainty activations

In many experimental and theoretical conceptualizations of uncertainty, prior and likelihood are not distinguished. However, as shown in this paper, a comparison of prior and likelihood uncertainty yielded quite different profiles: while prior uncertainty is largely associated with widespread activations in limbic and paralimbic brain areas and the putamen, likelihood uncertainty is associated with much more localized brain areas in the early stages of the dorsal sensory-motor pathway (see Figure 2.3 and Figure 2.5). Moreover, the effects of prior and likelihood are selectively different across these regions: The visual/superior occipital areas show an effect of likelihood uncertainty significantly higher than the effect observed in the putamen, amygdala, insula and OFC, which, by their turn, show an effect of prior uncertainty that is significantly higher than the effect observed in the superior occipital areas ($T>2$, $p<0.05$, paired t-test). That no overlap was observed between areas activated by prior and likelihood uncertainty and that direct comparison of the effects shows significant differences in activity patterns between the areas suggests that these forms of uncertainty are differentially represented in the nervous system

**Brain regions activated in successful trials**

To understand if scoring a point was rewarding to the subjects, we compared successful trials (in which subjects accurately guessed the position of the coin, and received a point as reward) with unsuccessful trials. In other words, if subjects derive a sense of reward for earning a point, then successful trials should activate reward associated areas. We find that successful trials were associated with strong bilateral activations in areas typically related with reward, including dorsal and ventral striatum and amygdala (Figure 2.8). There were essentially no areas that were significantly more active with unsuccessful vs. successful trials (at a threshold of $p<0.001$ unc.). The fact that we do not see significant activations upon unsuccessful trials further reinforces the idea that the activations we found related to prior or likelihood uncertainty cannot be explained by lower reward expectations. The fact that we see reward-related areas active upon success adds evidence that receiving a point is rewarding.

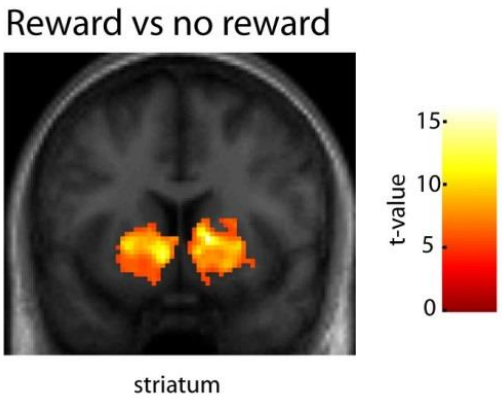


**Figure 2.8- Brain regions more activated in successful vs. unsuccessful trials.** Stronger activations were seen bilaterally in the striatum (y=8). Functional activations are overlaid on the coronal section of the subject-averaged anatomical scan (display threshold at $p<0.0001$ uncorrected, minimum 10 voxels).

**Behavioral data during the first (lab) session**

Variations of prior and likelihood uncertainty in our task significantly influenced the estimation behavior also if we looked at the behavior from all 27 subjects during the first behavioral session (performed outside the scanner). Subjects relied more on the likelihood information as the prior uncertainty increased and as the likelihood uncertainty decreased (main effect of prior, $F_{1,79}$= 173.74, $p<10^{-6}$; main effect of likelihood, $F_{1,79}$=45.43, $p<10^{-6}$, ANOVA repeated measures; see Figure 2.9-A). This shows that subjects could detect and react accordingly to variations in both prior and likelihood uncertainty. Furthermore, they did so in a way qualitatively predicted by Bayesian statistics.

Bayesian statistics does not tell us only that prior and likelihood information should be combined according to their uncertainties, it can also give us the optimal values given the associated uncertainties (see *2.3 Materials and methods*). The values obtained were qualitatively similar to what would be expected if people were using a Bayesian strategy (see Figure 2.9-A). Quantitatively, they are identical to the Bayesian optimal ones for the pl and PL conditions. The values for the pL and Pl conditions go in the "correct trend", but are statistically different from the optimal values ($T>2$, $p<0.05$, t-test). This might be due either to an overestimation of the narrow prior uncertainty and underestimation of the narrow likelihood uncertainty, or to an underestimation of both wide prior and likelihood uncertainty. The reasons for the difference with the optimal Bayesian values might indicate that subjects are using some heuristics that looks like Bayesian but is not fully so, and/or that we couldn't capture the "real" priors and likelihood uncertainties sensed by the subjects, which can be different from the experimentally imposed ones (due to sampling noise, noise in the visual system, etc.).

From the linear regression we can also obtain the mean of the prior that the subjects were using (see *2.3 Materials and methods* for details). We can see that, for all conditions, the mean that subjects used was identical to the real mean of the prior ($\mu_P$=0.5; see Figure 2.9-B), showing that subjects readily learned and used the information given about the prior's mean.

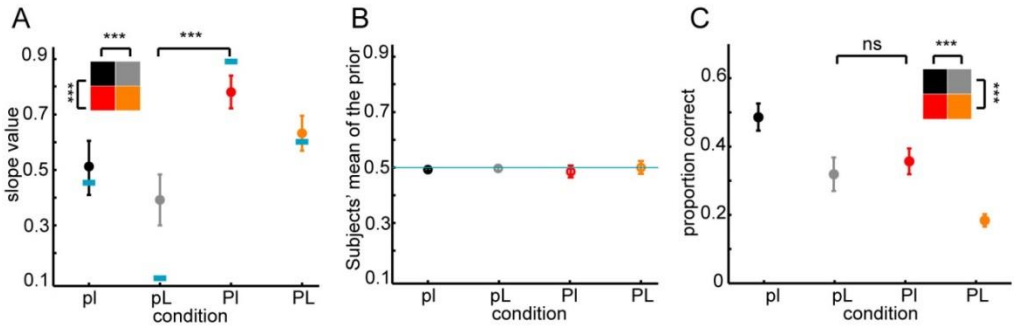


**Figure 2.9 - Behavioral results in the first session.** (A) Average slope of the linear regression for the behavior of the 27 subjects during the first behavioral session, separated by condition. The slope quantifies the degree to which subjects rely on the current visual stimulus (likelihood) vs. the prior. The small blue rectangles represent the optimal Bayesian values (see *2.3 Materials and methods* for details). (B) Average value for subjects' mean of the prior (see *2.3 Materials and methods* for details). The blue line represents the experimentally imposed mean of the prior ($\mu_P$=0.5). (C) Average proportion of trials in which the subjects accurately guessed the position of the target, separated by condition. Error bars in *(A). (B)* and *(C)* represent the 95% confidence interval for the mean. Inset in (A) and (C) shows that there is a significant effect of both prior and likelihood uncertainty. *ns*= non-significant, p>0.05; *** significant, p<0.001.

**Overlap between wide and narrow likelihood groups**

For the likelihood, two main conditions existed: a wide (more uncertain) likelihood, in which the dots were more spread apart ($\sigma_L$ = 0.15), and a more narrow likelihood ($\sigma_l$ = 0.06). Looking at the experimental standard deviations (std) we can see that the average for the two likelihood groups is indeed significantly different, and near the imposed std ($T=18.3$, $p<10^{-6}$, t-test, see Figure 2.10-A). Nevertheless, the between-trial variability in dot position creates overlap between the high and low uncertainty groups (see Figure 2.10-B). More specifically, we found that with the actual noise samples produced in our study 23% of the high uncertainty trials were not significantly different from 95% of the low uncertainty trials and, equivalently, 84% of the low uncertainty trials had dot dispersions that were indistinguishable from 95% of the high uncertainty trials. Hence, for more than half of the trials (54% on average), the std generated in one condition was within a 95% confidence interval for the stds generated by the other likelihood condition. Therefore, a parametric model that incorporates the specific sensory uncertainty per trial promises to give a more sensitive method to identify regions that track likelihood uncertainty (see Figure 2.5).

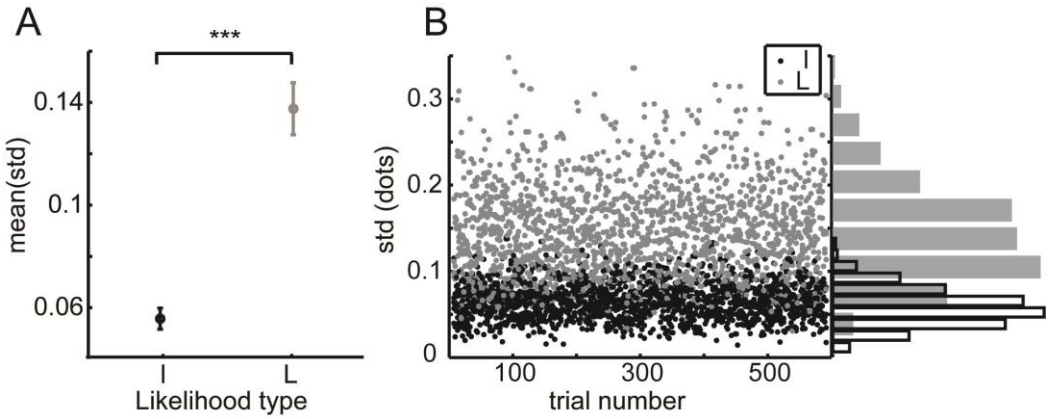


**Figure 2.10 - Overlap between wide and narrow likelihood groups.** (A) Average value for the standard deviations (std) of the blue likelihood dots during the narrow likelihood (l) and wide likelihood (L) conditions, for the 15 subjects during the fMRI experiment. Error bars represent the 95% confidence interval for the mean. *** significant, $p<0.001$. (B) Graphic representing the std value at each trial, for all the 15 subjects during the fMRI experiment. Black dots represent the std value of a particular narrow likelihood trial (l), and grey dots represent wide likelihood trials (L). Right to the graphic is a histogram of the corresponding data (black for l, grey for L).

## Controls

### Activity related to prior uncertainty is independent from reward expectation

Interestingly, uncertainty in prior information evokes fMRI activity in many of the same brain areas traditionally associated with reward processing (Cromwell and Schultz, 2003, Gottfried et al., 2003). This raises the possibility that we might have observed responses in these regions because prior uncertainty changes the expected or received reward. However, our two-by-two experimental design allowed us to compare two conditions (wide prior/narrow likelihood [Pl] vs. narrow prior/wide likelihood [pL]) that were matched in performance and expected reward (Figure 2.2-C), yet differed in the origin of the most relevant information (Figure 2.2-B). Contrasting these two conditions (Pl>pL) we observed BOLD activity increases across the same network of brain areas that appear in the high vs. low prior uncertainty contrast, including putamen, amygdala, insula and OFC ($p<0.001$ uncorrected (unc.)), implying that the increased activations arising as a consequence of prior uncertainty cannot be attributed merely to a lower reward expectation.

Furthermore, we performed an additional control analysis by incorporating each subject's performance in the fMRI model as a nuisance covariate, in this way modeling out performance/reward effects. This analysis again revealed significant activations in the same brain areas ($p<0.05$ FWE corrected), reinforcing the idea that the differential activations observed in putamen, amygdala, insula, and OFC were not related to different reward expectations.

**Activity is not related to posterior uncertainty or reaction times**

Given that prior and likelihood uncertainty together give rise to general (posterior) uncertainty, it is possible that some of the observed activations relate to posterior uncertainty. We can identify such effects using the diagonal comparisons, as behavioral performance should indicate the level of posterior uncertainty. For example, if the prior-related areas (Figure 2.3) specifically reflect prior uncertainty and not posterior uncertainty, then they should be present in the diagonal contrast Pl vs pL (different prior uncertainties but similar posterior uncertainties), but not in the contrast PL vs Pl or in the contrast pL vs pl (similar prior uncertainties but different posterior uncertainties). Direct comparisons of the beta values reveals that putamen, amygdala, insula, and OFC are significantly more active in the Pl vs pL contrast than in the PL vs Pl contrast or in the pL vs pl contrast ($p<0.05$, $T>1.76$, paired t-test), suggesting that the representations identified here with prior uncertainty are not associated with posterior uncertainty.

Similarly, we can ask if the results apparently related with likelihood uncertainty might be instead related with posterior uncertainty. Following the same logic as with prior uncertainty, we see that the superior occipital visual areas seen in the high likelihood uncertainty contrast are also significantly active in the diagonal comparison pL vs Pl (different likelihood uncertainties but similar posterior uncertainties), but not in the PL vs pL or Pl vs pl contrasts (different posterior uncertainties but similar likelihood uncertainties). Direct comparisons of the beta values reveals that the same superior occipital areas are significantly more active in the pL vs Pl contrast than in the PL vs pL or the Pl vs pl contrasts ($p<0.05$, $T>1.76$, paired t-test). Therefore, the increased activations observed in these superior occipital areas seem to relate specifically to likelihood uncertainty.

**Control task**

Although our task design allowed us to control for confounding effects of reward expectation and posterior uncertainty, there are several other variables that could have potentially altered interpretation of our results. For example, visual characteristics can differ across the four conditions in terms of the eccentricity of the observed dots, i.e. because the dots are more spread apart this might induce a broader, more dispersed area of foveal stimulation. Of course the fact that the dots are more spread apart indicates in itself that the visual stimulus is more uncertain, but we wanted to know if the same visual areas would appear even if the stimulus of interest has no uncertainty. Moreover, the four conditions differ with regard to the distance that subjects need to move the "net" on the screen. Together these issues raise the possibility that the observed fMRI activations could be driven by visual and motor features, independently of uncertainty. Therefore, in a control fMRI study that took place following the main experiment, subjects participated in an identical task, but in this case the target was revealed at the onset of the trial. In this way, the sensory input observed (visual dot dispersion) and the required motor output (movement to the predicted position) were matched to the main experiment, the only difference being that there was no uncertainty involved because the target was always visible.

With respect to the prior (high vs. low, same contrast as for Figure 2.3), we found a very different pattern from the one observed in the main experiment. The only significant areas ($p<0.05$, FWE at cluster level) consisted of activations of the left post-central gyrus and supplementary motor area (see Figure 2.11-A). Activations in these areas may be related to the increased average distance that the cursor was moved, and the fact that all subjects used their right hand to make the motor response. Critically, no significant activity was identified at the key regions identified from the main experiment, including putamen, amygdala, insula, and OFC, even at a

lenient statistical threshold (threshold at p=0.01 unc.). The interaction between prior uncertainty level (wide vs. narrow) and experiment (main study vs. control study) revealed these areas to be significantly more active in the main experiment than in the control (p<0.01 unc.). Together, these results indicate that the principal fMRI activations identified as a function of increased prior uncertainty (see Figure 2.3) were unlikely to be due to task-related sensory or motor confounds.

With respect to the standard analysis of the likelihood (high vs. low; same contrast as for Figure 2.5-A:B), fMRI activations were also observed in the occipital cortex, but centered medially around the calcarine sulcus, likely corresponding to visual cortical area V1 (Figure 2.11-B). These responses are approximately 15 mm away from the more dorsal superior occipital areas identified in the main contrast of likelihood uncertainty, suggesting once again that mere visual perceptual aspects of the task did not influence the main findings. Also, comparing directly the effects of likelihood uncertainty (wide vs. narrow) in the main experiment vs. the control we observe that the superior occipital areas are significantly more active in the main experiment than in the control (p<0.05 unc.). For the parametric analysis we also found largely non-overlapping areas (except for an area around the calcarine sulcus), and a direct comparison between trials in the main experiment and the control experiment demonstrated that these superior occipital areas are significantly more active during the main experiment (p<0.05, unc.). It thus seems that the dot dispersion *per se* does create a differential response in V1 (maybe due to different foveal stimulations), but that the stronger activations in the superior occipital cortex occur if the uncertain stimulus (with higher dot dispersion) is the stimulus of interest. Taken together, these controls suggest that the activations obtained with increased likelihood uncertainty are not merely due to different foveal representations.

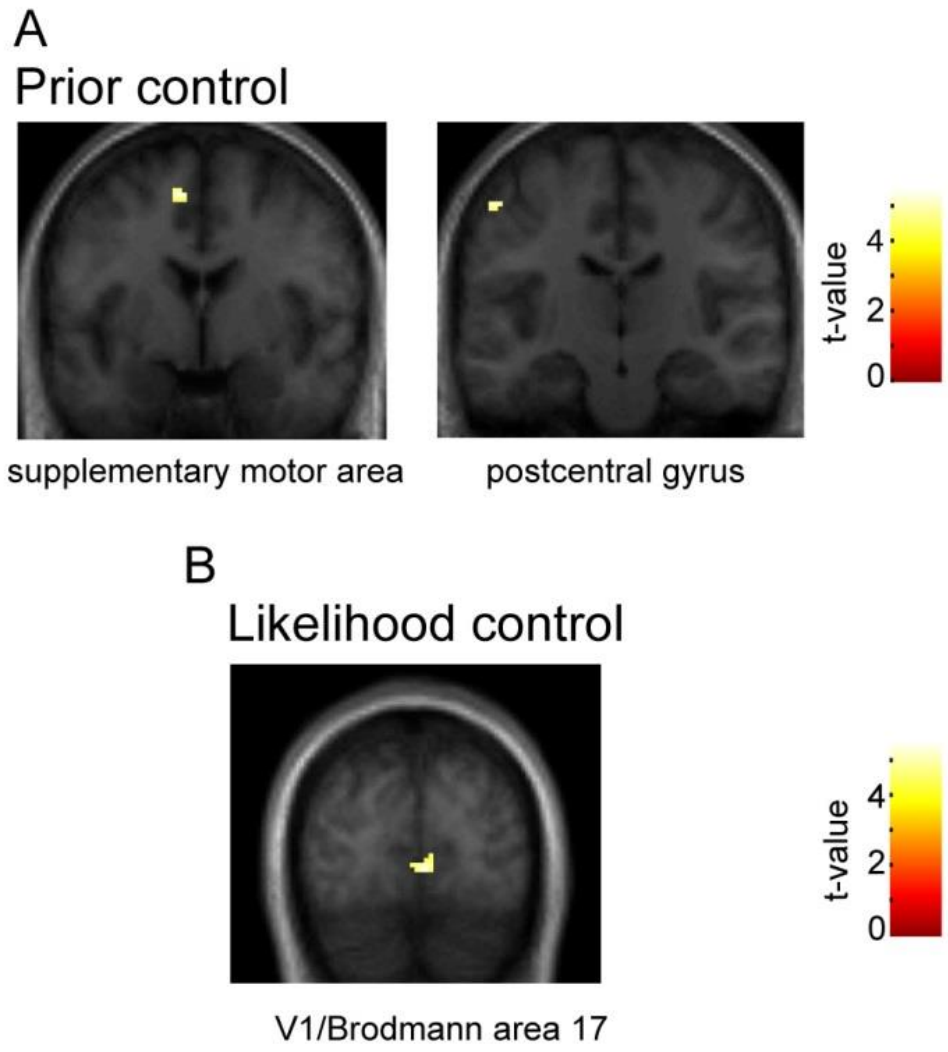


**Figure 2.11 - Brain regions activated during the control task.** (A) Stronger activations associated with the control block-equivalent high (vs. low) prior uncertainty were seen in the left supplementary motor area (y= -2) and post-central gyrus (y=-17). In this and all subsequent figures, the right side of the brain corresponds to the right side of the image. (B) Stronger activations associated with the control block-equivalent high (vs. low) likelihood uncertainty were seen in an area around the calcarine sulcus, roughly corresponding to V1 (y = -84). Functional activations are overlaid on the coronal section of the average subjects' T1-weighted scan (display threshold at p<0.0001 uncorrected, minimum 10 voxels).

## 2.5 Discussion

In this study we tested where prior and likelihood uncertainty are represented in the brain. We were particularly interested in knowing: 1) Are prior and likelihood uncertainty represented in the same set of areas? 2) Is uncertainty for visuomotor tasks represented in the traditional visuomotor pathway or in specialized areas? By combining a psychophysical paradigm with fMRI analysis in which two of the conditions were matched for performance and posterior uncertainty, we could disentangle the specific effects of prior and likelihood uncertainty. We found that greater prior uncertainty evoked increased brain activity in specialized brain areas that include the putamen, amygdala, OFC and parts of the insula. In contrast, greater likelihood uncertainty primarily affected neural activities in the occipital cortex, in areas that belong to the traditional visuomotor pathway. Prior and likelihood uncertainty were thus represented in largely non-overlapping areas, highlighting the importance of distinguishing between these kinds of uncertainty.

There are numerous reasons why the brain could use a strategy where prior uncertainty is processed in these specialized brain areas. Computationally, the forming of a prior results from integration over time, requiring long-term memory, and may be more difficult to implement in domain-specific sensory/motor brain areas. Anatomically, OFC, amygdala, putamen and insula receive inputs from sensory areas, which are needed to build a prior, and project to motor effector systems in the brainstem and cortex(Albin et al., 1989, Whalen, 1998, Rolls, 2000, Singer et al., 2009) which are needed to use the prior. These connections may allow the optimization of behavior in the context of varying prior uncertainty and, if the prior is too uncertain, facilitate behavior change to gather relevant new information(Whalen, 1998). Additionally, these areas are connected or even directly involved with reward processing (Rolls, 2000, Preuschoff et al., 2006, Rangel et al., 2008), and could thus more easily combine the need for

more information with its potential value. Moreover, previous research has highlighted their involvement in uncertainty in reward(Critchley et al., 2001, Huettel et al., 2005, Tobler et al., 2007, Kepecs et al., 2008, Singer et al., 2009) and, more specifically, in signaling ambiguity and the need to learn more about the world(Hsu et al., 2005). Thus, computational demands, anatomical connections, and previous research support the involvement of these areas in signaling prior uncertainty.

Likelihood uncertainty, in contrast, was associated with activations in areas that are part of the traditional visuomotor pathway. Our task is visual, and thus information about the task is also transmitted through the same route. Hence, these results concur with the hypothesis presented by theoretical models, such as probabilistic population codes and sampling theories(Ma et al., 2006, Fiser et al., 2010), that likelihood uncertainty is part of the inherent code by which neurons transmit information. According to these models, the activity of the same neurons transmits information (e.g. the position) along with uncertainty about this information. From a computational perspective, sensory information needs to be continuously used to calculate estimates of likelihood uncertainty(Burge et al., 2008) and hence sensory areas seem best suited for this ongoing update. Indeed, previous studies in human visual perception have also found that uncertainty affects brain activity in the corresponding sensory areas(Beauchamp et al., 2010, Helbig et al., 2012), and, moreover, they found that activity in visual areas was higher when the visual stimulus was more uncertain(McKeefry et al., 1997, Murray et al., 2002, Beauchamp et al., 2010). This occurred for random/ nonrandom dot motion(McKeefry et al., 1997), incoherent/coherent shapes(Murray et al., 2002) and blurry/non-blurry images(Helbig et al., 2012), indicating that the effects observed are not exclusive to the particular stimulus we used. Together, our results suggest that likelihood uncertainty about a visual stimulus may be processed along with the stimulus itself in the visual cortex.

For Bayesian decision-making, the brain needs not only to compute prior and likelihood uncertainty, but also to use them for appropriate weighting of both pieces of information. Activity in the brain areas where this occurs should then relate to these weights, which depend nonlinearly on both uncertainties. We found that an area in the PFC tracks the trial-by-trial weight on current vs. prior information. This area would then be a candidate area to receive information from both prior and likelihood uncertainty and calculate accordingly how much weight should be placed on new information. Indeed, the PFC is known for its role in planning and cognitive control(Miller, 2000) and, interestingly, has even been specifically associated with Bayesian decision-making(Hampton et al., 2006). Our results thus suggest how a network of brain areas may give rise to Bayesian instantiations of perception and behavior.

Although our study focused on the integration of priors with visual information, uncertainty may be represented differently for other sensory modalities or tasks. Our task dealt with a new, rapidly acquired prior over a series of trials, in which subjects, during each trial, had unlimited time to make a decision. Hence, it can be considered a cognitive task. Studies using different kinds of tasks, but that were also cognitive, have previously associated the prior-uncertainty areas identified here with decision-making and reward uncertainty(Critchley et al., 2001, Huettel et al., 2005, Preuschoff et al., 2006, Brand et al., 2007), making it more likely that our results would hold for other types of cognitive tasks as well. However, it is possible that other types of priors, such as the ones involved in early sensory perception (e.g. “light from above”), which might have been learned over many years (or even generations), have distinct representations(Fischer and Pena, 2011). As for likelihood uncertainty, if indeed it is transmitted concurrently with the sensory information itself, then a non-visual task should activate non-visual brain areas. Interesting future studies could help to answer these questions, for example by applying a

computationally equivalent task in which the sensory feedback information given is not visual but, say, olfactory or auditory; or by implementing a faster/more unconscious sensory perception task.

The association of specific brain areas with prior uncertainty and potentially with subjects' individual tendency to rely on new sensory information (as is the case for putamen) may have implications for the understanding of learning disabilities and abnormal decision-making behavior. Learning, in a Bayesian sense, can be interpreted as the weight given to new evidence over prior beliefs(Courville et al., 2006). If the prior is more certain then less learning from new information should occur, and vice versa(Courville et al., 2006). Changes in the brain areas that represent prior uncertainty might then lead for example to an underestimation of prior uncertainty, potentially affecting learning from new information. Interestingly, it has been found that patients with damage to the amygdala(Brand et al., 2007), OFC(Hsu et al., 2005), insula(Clark et al., 2008) and putamen(Shohamy et al., 2004) show considerable deficits in making decisions that involve uncertainty and learning from feedback, and that the existence of intact connections between these structures is essential for learning from new sensory information(Baxter et al., 2000, Cohen et al., 2008). If indeed one of the reasons why these deficits occur is based on an underestimation of prior uncertainty, then these could potentially be reversed by providing more certain current sensory information or by giving explicit information about prior uncertainty.

The finding of the areas involved in prior and likelihood uncertainty representation provides insight for an ongoing debate in the computational literature: is uncertainty part of the general code by which neurons exchange information and thus encoded in every neuron's output(Ma et al., 2006, Fiser et al., 2010), or are there specialized areas that deal with the encoding of uncertainty(Hsu et al., 2005, Preuschoff et al., 2006, Schultz et al., 2008, Singer et al., 2009)? Our findings suggest that both of these

hypotheses might be correct, but for different kinds of uncertainty: likelihood uncertainty seems intrinsically embodied in the stimulus encoding itself, as it is represented in sensorimotor areas, while prior uncertainty is encoded in specialized areas. Future models of brain function should take into account uncertainty in the prior and in the likelihood separately, since both their neural representations and their behavioral effects are distinct.

### 2.6 Acknowledgements

This work was supported by PhD grants SFRH/BD/33272/2007 (to I.V.), SFRH/BD/33525/2008 (to H.L.F.), NIH grants 2P01NS044393 and 1R01NS063399 (to K.P.K.), NIH grants 1R01DC010014 and K08DC007653 (to J.A.G.), and a Northwestern F32 Institutional Training Grant (to J.D.H.). We would like to thank Peter Dayan for helpful comments on the manuscript.

### 2.6 References

Albin RL, Young AB, Penney JB (1989) The Functional-Anatomy of Basal Ganglia Disorders. Trends Neurosci 12:366-375.

Basso MA, Wurtz RH (1997) Modulation of neuronal activity by target uncertainty. Nature 389:66-69.

Baxter MG, Parker A, Lindner CCC, Izquierdo AD, Murray EA (2000) Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. J Neurosci 20:4311-4319.

Beauchamp MS, Pasalar S, Ro T (2010) Neural substrates of reliability-weighted visual-tactile multisensory integration. Front Syst Neurosci 4:25.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nature Neuroscience 10:1214-1221.

Berniker M, Voss M, Kording K, Brezina V (2010) Learning priors for Bayesian computations in the nervous system. Plos One 5:e12686.

Brand M, Grabenhorst F, Starcke K, Vandekerckhove MMP, Markowitsch HJ (2007) Role of the amygdala in decisions under ambiguity and decisions under risk: Evidence from patients with Urbach-Wiethe disease. Neuropsychologia 45:1305-1317.

Burge J, Ernst MO, Banks MS (2008) The statistical determinants of adaptation rate in human reaching. Journal of Vision 8.

Clark L, Bechara A, Damasio H, Aitken MRF, Sahakian BJ, Robbins TW (2008) Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. Brain 131:1311-1322.

Cohen MX, Elger CE, Weber B (2008) Amygdala tractography predicts functional connectivity and learning during feedback-guided decision-making. Neuroimage 39:1396-1407.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences 10:294-300.

Cox RT (1946) Probability, Frequency and Reasonable Expectation. American Journal of Physics 14:1-13.

Critchley HD, Mathias CJ, Dolan RJ (2001) Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron 29:537-545.

Cromwell HC, Schultz W (2003) Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. Journal of Neurophysiology 89:2823-2838.

Duvernoy HM (1999) The human brain: surface, blood supply and three-dimensional sectional anatomy: Springer Verlag Wien.

Fischer BJ, Pena JL (2011) Owl's behavior and neural representation predicted by Bayesian inference. Nature Neuroscience 14:1061-U1163.

Fiser J, Berkes P, Orban G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14:119-130.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 2:189-210.

Glascher J, Daw N, Dayan P, O'Doherty JP (2010) States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron 66:585-595.

Glimcher PW (2003) Decisions, uncertainty, and the brain: The science of neuroeconomics. Cambridge, MA: MIT Press.

Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Sciences 5:10-16.

Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104-1107.

Graves LE, Ridgers ND, Williams K, Stratton G, Atkinson G, Cable NT (2010) The physiological cost and enjoyment of Wii Fit in adolescents, young adults, and older adults. J Phys Act Health 7:393-401.

Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci 26:8360-8367.

Helbig HB, Ernst MO, Ricciardi E, Pietrini P, Thielscher A, Mayer KM, Schultz J, Noppeney U (2012) The neural mechanisms of reliability weighted integration of shape information from vision and touch. Neuroimage 60:1063-1072.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Huettel SA, Song AW, McCarthy G (2005) Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci 25:3304-3311.

Jazayeri M, Shadlen MN (2010) Temporal context calibrates interval timing. Nat Neurosci 13:1020-1026.

Josephs O, Henson RNA (1999) Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philosophical Transactions of the Royal Society B-Biological Sciences 354:1215-1228.

Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. Nature 455:227-U255.

Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244-247.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432-1438.

McKeefry DJ, Watson JDG, Frackowiak RSJ, Fong K, Zeki S (1997) The activity in human areas V1/V2, V3, and V5 during the perception of coherent and incoherent motion. Neuroimage 5:1-12.

Miller EK (2000) The prefrontal cortex and cognitive control. Nat Rev Neurosci 1:59-65.

Miyazaki M, Nozaki D, Nakajima Y (2005) Testing Bayesian models of human coincidence timing. Journal of Neurophysiology 94:395-399.

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002) Shape perception reduces activity in human primary visual cortex. P Natl Acad Sci USA 99:15164-15169.

Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. Neuron 51:381-390.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545-556.

Rolls ET (2000) The orbitofrontal cortex and reward. Cerebral Cortex 10:284-294.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P (2008) Explicit neural signals reflecting reward

uncertainty. Philosophical Transactions of the Royal Society B-Biological Sciences 363:3801-3811.

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of Neurophysiology 86:1916-1936.

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA (2004) Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 127:851-859.

Singer T, Critchley HD, Preuschoff K (2009) A common role of insula in feelings, empathy and uncertainty. Trends in Cognitive Sciences 13:334-340.

Tassinari H, Hudson TE, Landy MS (2006) Combining priors and noisy visual cues in a rapid pointing task. J Neurosci 26:10154-10163.

Tobler PN, O'Doherty JP, Dolan RJ, Schultz W (2007) Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. Journal of Neurophysiology 97:1621-1632.

Vilares I, Kording K (2011) Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Year in Cognitive Neuroscience 1224:22-39.

Whalen PJ (1998) Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. Curr Dir Psychol Sci 7:177-188.

# 3. DECISION-MAKING UNDER UNCERTAINTY IN PARKINSON'S DISEASE

# 3. DECISION-MAKING UNDER UNCERTAINTY IN PARKINSON'S DISEASE

**Citation:** Vilares I and Kording KP (*under review*). Decision-making under uncertainty in Parkinson's Disease.

**Author Contributions:** All authors contributed in designing the study. IV performed the experiments; IV analyzed the data; IV and KK wrote the paper.

## 3.1 Summary

Dopamine and putamen activity are crucial for decision-making under uncertainty. However, their specific role is still a subject of intense debate. To test potential roles, we had Parkinson's disease (PD) patients do a visual decision-making task in which both prior and current sensory uncertainty (likelihood) were varied and where behavior is often predicted by Bayesian statistics. We found that many aspects of uncertainty processing were conserved in PD: both groups could learn prior distributions and utilize priors and current sensory information. However, in the PD group, in particular when off drugs, subjects showed a much weaker sensitivity to differences in likelihood uncertainty. Our results suggest that dopamine and putamen activity, which are affected by PD, have a crucial role in the processing of current sensory uncertainty.

*"Se podes olhar, vê. Se podes ver, repara".*

*"If you can look, see. If you can see, notice".*

*José Saramago*

### 3.2 Introduction

Every day we are faced with decisions that have associated uncertainty. They may range from very important, life-changing situations (should I marry this person? Which career should I choose?), to minor decisions that we make, sometimes without even noticing that we are doing them (where is my wallet? What is that object in the sky?). When making a decision, we combine information from the past, called prior (e.g. how trustworthy has this person shown to be? What types of objects generally appear in the sky?), with current sensory information, or likelihood (what am I feeling at this time? What is the shape of the object that I am seeing now?). Correctly combining these pieces of information is the key to effective decision-making.

Bayesian theory tells us that the best way to combine these pieces of information is to join them according to their respective uncertainties. Hence, understanding how the brain performs these calculations seems of extreme importance if we want to understand what goes wrong when there is faulty decision-making. Previous research in the lab has indicated that the putamen was particularly important in this process, with activity from the putamen correlating with both increased prior uncertainty but also with individual's tendencies to sense and attend to current vs. prior information (Vilares et al., 2012). However, from the fMRI data alone it is not possible to know in which of these processes putamen activity has a causal role. Better understanding the role of putamen may shed light on how the brain performs the computations necessary for correct decision-making under uncertainty.

Existing literature shows that activity from the putamen either directly or indirectly affects decision-making under uncertainty. For example, putamen activity has been associated with ambiguity, i.e. with lack of knowledge of the real probabilities of a given event (Hsu et al., 2005). It has also been implicated in learning (Grafton et al., 1995, Lehericy et al., 2005, Poldrack et al., 2005, Doyon et al., 2009, Orban et al., 2010). Interestingly, when learning a motor sequence putamen activity decreases after training, but remains high if the sequence is random (Lehericy et al., 2005, Poldrack et al., 2005). These studies can be interpreted as putamen activity signaling for prior uncertainty, which decreases with learning but keeps being high if the sequence is random or the probabilities of events are unknown. Another interpretation is that it allows focusing on the current incoming information, which would equally predict the observed data. As activity from the putamen is severely compromised in people that suffer from Parkinson's disease (Kish et al., 1988, Playford et al., 1992) we can use Parkinson's disease (PD) patients to ask fundamental questions about the role of the putamen in decision-making under uncertainty.

PD leads to a depletion of dopamine (Kish et al., 1988) which probably mediates the putamen's influence on decision-making under uncertainty. Indeed, dopamine has been associated with uncertainty (Fiorillo et al., 2003, Friston, 2009, Friston et al., 2012). However, its precise role is still unclear and subject to intense debate (Berridge and Robinson, 1998, Wise, 2004, Friston et al., 2012). It has been proposed that dopamine is involved in attention and saliency of a stimulus (Berridge and Robinson, 1998, Kapur, 2003, McClure et al., 2003), even when that stimulus does not predict reward (Horvitz, 2000). It has also been proposed that dopamine regulates the weight given to bottom-up current sensory information (likelihood) vs. top-down prior beliefs (Friston, 2009, Beeler et al., 2010, Friston et al., 2012, Galea et al., 2012), and that it codes for the uncertainty in the current stimulus (Friston, 2009, Friston et al., 2012, Galea et al., 2012).

Furthermore, dopamine has been implicated in learning (Dayan and Balleine, 2002, McClure et al., 2003, Frank et al., 2004), and hence could be crucial for learning the uncertainty of a given event. To distinguish between these ideas we can compare PD patients with healthy subjects on a task that allows disambiguating different aspects of decision-making under uncertainty (Vilares et al., 2012).

Previous studies have shown that PD patients have impairments in decision-making. They have trouble at reacting to unexpected events (Galea et al., 2012), task switching (Cools et al., 2003) and shifting mental sets according to external cues (Mimura et al., 2006). They are also impaired in making decisions under ambiguity (Delazer et al., 2009) and in using negative feedback (Brand et al., 2004). It seems reasonable to assume that some of these deficits may actually result from the wrong processing of uncertainty information, e.g. an overreliance on prior experience or a diminished capacity to attend external cues. Studying how PD patients learn and deal with uncertainty in prior and current sensory information may help in the understanding of which specific challenges PD patients have to face when making decisions under uncertainty.

Here, we had Parkinson disease patients and age-matched controls perform a decision-making task in which they could use both prior and likelihood to estimate the position of a target. Patients performed the task once “on” medication (more dopamine in their system) and once “off” medication (less dopamine). Prior and likelihood uncertainty were varied independently in a two-by-two factorial design, and for each condition the Bayesian slope, i.e. the weight given to current vs. prior information, was obtained. In this way, we could independently analyze the capacity of learning the prior, the average weight given to likelihood vs. prior information, and the ability to attend and react to differences in likelihood uncertainty. Patients performed significantly worse in this task when compared to age-matched controls, and that performance could be ameliorated with dopamine administration.

Moreover, patients could learn the prior as well as controls and did not show significant differences on overall reliance on current vs. prior information. However, they were impaired at reacting to differences in likelihood uncertainty, and this impairment was particularly strong when they were *off* medication. Our results indicate that PD patients have deficits in reacting to differences in likelihood uncertainty, suggesting a causal role of dopaminergic transmission from the putamen in the processing of uncertainty.

### 3.3 Materials and methods

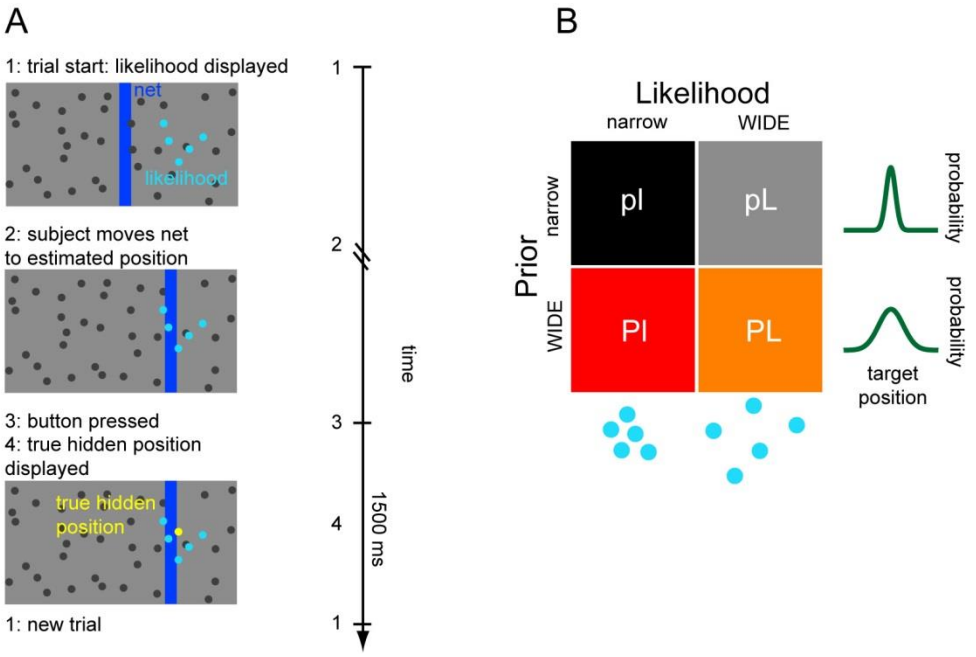


**Figure 3.1 - Experimental setup.** (A) Illustration of the task. Subjects guess the position of a hidden target (the "coin", represented by the yellow dot) using a net (vertical blue bar) which they can displace horizontally. At the onset of each trial, subjects receive noisy information about the position of the hidden target in the form of a set of 5 blue dots (the likelihood). Subjects then move the net to the guessed position and press the mouse button to confirm their choice, after which the true target position is displayed. A new trial then begins 1500 ms later. Left: illustration of the computer display that was presented to the subjects. Right: typical time course of a trial. (B) The 4 conditions of the experiment. The experiment consisted of a two-by-two factorial design, with two types of prior (p=narrow prior; P=wide prior) and two types of likelihood (l=narrow likelihood; L=wide likelihood). The wider conditions are the ones with more associated uncertainty.

**Subjects**

23 Parkinson disease (PD) patients and 16 healthy controls were recruited for the study. From these, 8 patients were discarded: 6 due to failure to follow the "on-off" medication schedule required for the experiment, 1 because she was still not taking meds and 1 because he couldn't understand the task even after several explanations. One healthy control was discarded because she had already participated in a very similar study in the lab before. Thus, in total, 15 PD patients (9 women) and 15 non-PD controls (8 women) participated in the experiment. Subjects were paid $20 for participating in each session of the experiment, so they got $40 for both sessions. Written informed consent was obtained for all subjects. All protocols were approved by the Northwestern University IRB.

PD patients: All the PD patients were recruited from the Rehabilitation Institute of Chicago and the Northwestern Memorial Hospital. All PD patients had idiopathic Parkinson's disease as diagnosed by a neurologist. The 15 patients included in the study had between 43 to 81 years old (62.8 ± 9.6 years old; mean ± std). Severity of disease was equivalent to Hoehn and Yahr stages ranging between I and III. The duration of Parkinson's disease varied from 1.5 months to 18 years from the time of initial diagnosis (6.2 ± 4.8 years; mean ± std). Eleven of the fifteen patients included in the study were taking daily L-dopa preparations. Of those, eight received additional pharmacotherapy with dopamine receptor agonists, rasagiline or amantadine. Four patients were taking dopamine receptor agonists and/or rasagiline and/or amantadine. L-dopa equivalent units (LEU) of patients' regular daily dopamine replacement therapies were calculated as described elsewhere (Tomlinson et al., 2010), and varied between 100 mg to 1733 mg (675 ± 436 mg; mean ± std).

Healthy controls: Age-matched healthy volunteers, with no current known neurological problems, were recruited for the study, usually from amongst

the patients' spouses or partners. The 15 healthy controls included in the study had between 43 to 84 years old (63.4 ± 11.6 years old; mean ± std). No significant difference in age existed between PD patients and Controls (*p=0.88*, t-test).

**General procedure**

PD patients completed two test sessions, once about 1 hour after they had taken their regular dopaminergic medication (*on-state*) and once after overnight withdrawal from dopaminergic medication (*off-state*). Session order was counterbalanced across patients, so that in total 8 patients started in the *on-state*, and 7 started in the *off-state*. Control subjects also completed two test sessions, and the results reported for controls were the average of the results obtained in the two sessions. For both PD patients and controls, the two sessions were generally done in two consecutive days.

**Coin-catching task**

Subjects performed a decision-making task in which they had to guess the position of a hidden coin on a screen (Vilares et al., 2012). They were told the cover story of a coin being tossed into a pond and informed that their task was to guess where the coin had fallen. They could not see the coin, but they could see 5 blue dots that were the "splashes" produced by the coin falling in. They were told that the person who threw the coin aimed, albeit imperfectly, at the center of the screen (mean of prior). They were also told that, between blocks, the thrower changed, and the new one might be better or worse at throwing (what indirectly informed them that the variance of the prior changed). To estimate the coin position, subjects could use (although they were never explicitly told so) both the coin position's likelihood,

obtained from the “splashes”, and its prior (the distribution of previous coin locations). There was no temporal deadline. These instructions were scripted and given to the subject to read, to minimize variability in explanation. If something was not understood, the researcher would repeat the main points until the subject acknowledged understanding.

**Stimuli**

The position of the coin was drawn from a Gaussian distribution which was centered on the center of the screen, and with a standard deviation (std) that was either low ($\sigma_p$=2.5% of screen width) or high ($\sigma_P$ = 8.5% of screen width). This distribution was the *prior* of the experiment. Subjects were given the mean of the prior (“the coin throw is aimed at the screen center”) but not its variance, which they could only estimate from the distribution of previous coin throws. The standard deviation of the prior was kept constant within blocks, but changed across blocks. On every trial, a cluster of five dots was shown on the screen. The x-position of each of these dots was drawn independently from a second Gaussian distribution in which the mean was the coin’s horizontal location on that trial and the standard deviation was either low ($\sigma_l$ = 6%) or high ($\sigma_L$ = 15%). The distribution of these five dots defined the *likelihood*. The std of the likelihood was varied pseudo-randomly from trial to trial but counterbalanced across trials. We made the std of the likelihood vary pseudorandom from trial to trial so that subjects could not predict *a priori* the overall uncertainty that the trial would have. In total there were thus four conditions: low prior uncertainty and low likelihood uncertainty (**pl**); low prior uncertainty and high likelihood uncertainty (**pL**); high prior uncertainty and low likelihood uncertainty (**Pl**) and high prior uncertainty and high likelihood uncertainty (**PL**). The std values used were the same ones reported in (Vilares et al., 2012).

The colors and appearance of the stimulus were also identical to the ones reported in (Vilares et al., 2012) (see Figure 3.1-A). The screen units were normalized between 0 (the left edge) and 1 (the right edge). Stimulus presentation was performed using Matlab R2012a (MathWorks, Natick, MA).

**Task procedure**

Experimental blocks: At the onset of each trial, five blue likelihood dots were shown on the screen, where they remained until the end of the trial. Subjects had to move a blue vertical bar (the "net") to estimate the coin position. Contrary to the task in (Vilares et al., 2012), in order for the subjects to move the net they had to press either the "left" or the "right" keyboard key (which would move the net, respectively, to the left or the right). Once they made their decision, they would press the space bar. The change from the mouse to the keyboard was done to help subjects perform the task, given that preliminary data suggested that older subjects had troubles using the mouse. Subjects could take as long as they wanted to decide where to place the net but had to wait for at least 1.26 seconds (to minimize accidental double presses counting as a response). After they pressed the button, the true position of the coin was revealed and subjects would get one point added to the score if the coin was inside the net. The cumulative score across the experiment was shown to the subjects at the end of each trial. A new trial would then begin 1.5 seconds later (see Figure 3.1-A). Given that the net covered the entire height of the screen only the horizontal location was relevant, making this a one-dimensional estimation task. Subjects completed two blocks of 150 trials each for a total of 300 trials per experiment. The first block had a low prior uncertainty and the second one had a high prior uncertainty. We had to decrease the total number of blocks performed (in comparison with (Vilares et al., 2012)) because preliminary data indicated that subjects were getting too tired and

unwilling to do 600 trials. After finishing the 2 blocks in the second session, subjects also performed 100 trials of a control block. Each trial took an average of 8.5 seconds and the total experiment lasted on average approximately one hour.

Control Block: The control block was performed immediately following completion of the two experimental blocks, on the second test session. The task was identical to the main experiment, with the only difference being that the coin's location was shown at the onset of each trial and could be seen throughout, so there was no uncertainty about its position. The 100 trials comprising the control block were selected by randomly sampling 25 trials per condition from the main experiment blocks. Each trial of the control block repeated one of those sampled experimental trials, showing the same likelihood dot display that was shown in the experimental trial and using as the coin position the actual position to which the person moved the net at that trial. As in the main experiment, subjects were awarded one point for successfully moving the net to the coin's position. This control block was done in order to understand if subjects were able to accurately move the net to the required place and press the button. All subjects included in the analysis were able to perform the control block without apparent problems.

**Data Analysis**

Bayesian modeling of behavior: In order to perform the task successfully, for every trial the subjects should place the net in the most likely location of the hidden coin. Bayes rule provides an optimal way to estimate this location(Kording and Wolpert, 2004):

$$X_{est}=\sigma^2_L/(\sigma^2_L+\sigma^2_P)\mu_P+\sigma^2_P/(\sigma^2_L+\sigma^2_P)\mu_L \qquad (1)$$

where $X_{est}$ is the estimated position of the coin, $\sigma^2_L$ and $\sigma^2_P$ are the variances of the likelihood and of the prior, respectively, and $\mu_L$ and $\mu_P$ are

the respective means. For our experiment, the real $\mu_P$ (the mean of the prior) is always a constant, in this case 0.5 (the center of the screen). The mean of the likelihood ($\mu_L$) for each particular trial can be considered the centroid of the cloud of dots, and it changes from trial to trial. We can then make a linear regression of the subject's estimated coin position, $X_{est}$, as a function of the centroid of the cloud of dots. The slope of this linear regression, the **Bayesian slope** (s), characterizes how much the subject is weighting the current sensory information (likelihood), and, if people perform according to the optimum prescribed by Bayesian statistics, its value should be equal to the perceived $\sigma^2_P/(\sigma^2_L+ \sigma^2_P)$. A slope of zero suggests that subjects do not take into account likelihood information, and a slope of one suggests that subjects only use likelihood information. A slope between zero and one indicates that subjects are using information from both prior and likelihood, and the larger the slope the more they rely on the likelihood and less on the prior.

Subjects' average reliance on likelihood vs. prior information: the slopes associated with each of the four conditions were averaged to get an estimate of an individual's average reliance on likelihood information (i.e. $(s_{pl}+ s_{pL}+s_{Pl} +s_{PL})/4$).

Subjects' performance: The proportion of correct trials, for each condition, was calculated as the number of trials in which the subject accurately guessed the position of the coin, divided by the total number of trials in that condition.

Sensitivity to prior uncertainty: Is the difference between the Bayesian slopes in one type of prior and the slopes in the other type of prior (i.e. $((s_{Pl} +s_{PL} )-(s_{pl}+ s_{pL}))/2$).

Sensitivity to likelihood uncertainty: Is the difference between the Bayesian slopes in one type of likelihood and the slopes in the other type of likelihood (i.e. $((s_{pI}+s_{PI})-(s_{pL}+s_{PL}))/2$).

For the behavioral calculations, the first 20 trials of every block were discarded to minimize the effect of learning.

Optimal values: The Bayesian optimal values for the slopes can be obtained from $\sigma^2_P/(\sigma^2_L+\sigma^2_P)$, where $\sigma^2_P$ is the variance associated with the prior (with $\sigma^2_p=0.025^2$ or $\sigma^2_P = 0.085^2$ in unit-less screen coordinates) and $\sigma^2_L$ is the variance associated with the likelihood, which can be estimated by $\sigma^2_L$= variance(cloud of dots)/ number of dots (so in our case, $\sigma^2_I=0.06^2/5$ or $\sigma^2_L = 0.15^2/5$).

### 3.4 Results

In this study we wanted to understand how Parkinson's disease (PD) patients learn and combine uncertain prior and current sensory information, and how it differs from controls. We recruited 15 PD patients and 15 age-matched controls to perform a visual decision-making task. In this task, subjects had to guess where a hidden target ("coin") would appear on a screen (see Figure 3.1 and (Vilares et al., 2012)). They received noisy visual information about the position of the target in the form of a dot-cloud which was centered at the true target position. To accurately estimate the position of the target, subjects could use both the prior, obtained from the distribution of previous target positions, and the likelihood, obtained from the displayed dots (see methods for details). The conditions comprised a two-by-two factorial design (Figure 3.1-B), with two levels of prior uncertainty (wide, more uncertain prior: P; and narrow, less uncertain prior: p) and two levels of likelihood uncertainty (wide likelihood: **L**; and narrow likelihood: **l**). Subjects performed the task twice, with PD patients performing it once "on" medication (shortly after taking their dopaminergic medication) and once "off" medication (dopaminergic medication withdrawn overnight). By varying prior and likelihood uncertainty, comparing healthy controls and patients and changing dopamine medication, we could study the specific effects of the disease and of dopamine in decision-making under uncertainty.

**PD patients perform worse than age-matched controls**

First we wanted to understand if PD affected performance in the coin-catching task. We found that, on average, patients (*off-state*) guessed where the hidden target was significantly fewer times than their age-matched controls ($p=0.03$, t-test, n=15; see Figure 3.2-A). Performing a 3-way repeated measures ANOVA (*population type x types of prior uncertainty x types of likelihood uncertainty*) we can see that, besides the significant main effect of population type, there was also a significant main effect of both prior and likelihood uncertainty on performance ($F_{(1,29)}>16.6$, $p<0.001$; see Figure 3.2-B), with higher levels of uncertainty leading to a decrease in performance. This is not surprising, given that both types of uncertainty may be expected to change the precision of subjects' estimates and thus their expected task performance. There were no significant interaction effects ($F_{(1,29)}<2.7$, $p>0.1$). The difference in performance between the PD patients and controls does not seem to be related simply to lower motor ability, as there was no significant difference in how much patients and controls moved the net from the starting point ($p=0.46$, t-test, n=15). PD impaired performance in a task that varies uncertainty in both prior and current information.

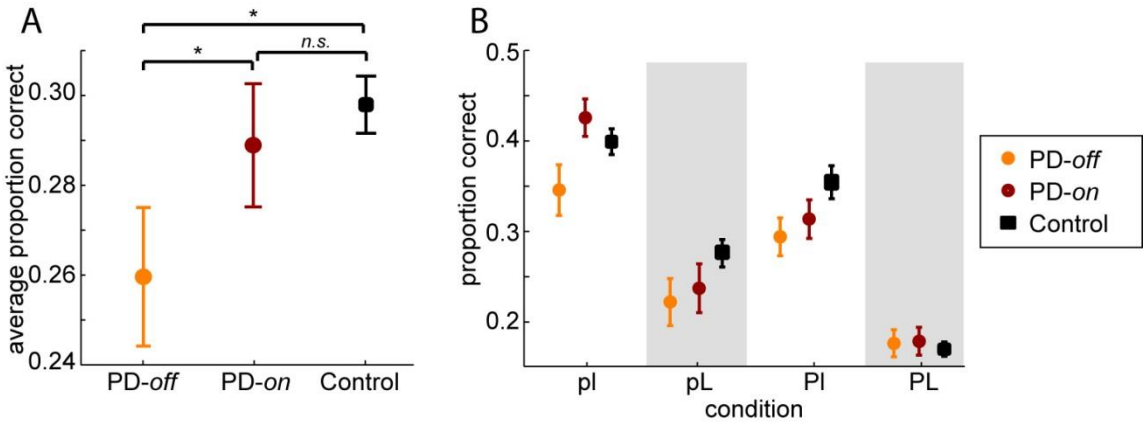


**Figure 3.2 - Performance in the task:** average proportion of trials in which subjects correctly guessed the position of the target. Average ± standard error of the mean (SEM) for PD patients *off-state* (orange circles), PD patients *on-state* (red circles) and controls (black squares), averaged across conditions (A) or separated by condition (B). On average, PD patients performed significantly worse when they were *off-state* compared with when they were *on-state* (*p=0.01*, paired t-test, n=15) or compared to age-matched controls (*p=0.03*, t-test, n=15).

### Dopamine improves performance

Next we wanted to understand how administration of dopaminergic replacement medication affects performance in the task. PD patients *on-state* (i.e. when they have more dopamine in their system) performed significantly better compared with when they were *off-state* (*p=0.01*, paired t-test, n=15; see Figure 3.2-A), showing performance levels similar to controls (*p=0.56*, t-test, n=15). A 3-way repeated measures ANOVA (*medication state x types of prior uncertainty x types of likelihood uncertainty*) showed, besides the significant main effect of medication, significant main effects of prior and likelihood uncertainty on performance ($F_{(1,57)}=7.8$, $p=0.015$ and $F_{(1,57)}=109$,$p<0.001$, ANOVA-rm; see Figure 3.2-B). There was also a significant interaction between the effects of medication and likelihood uncertainty ($F_{(1,57)}=4.5$,$p=0.038$). The other interactions did

not reach significance ($F_{(1,57)}<3.7$, $p>0.05$). Our results show that dopaminergic replacement medication improved performance in the task.

**No difference in the capacity of learning the prior uncertainty**

If PD affects the learning of the uncertainty in the prior, then we would expect PD patients (*off-state*) to not change behavior as much between blocks of different prior uncertainties, i.e. to be less sensitive to differences in prior uncertainty. The main behavioral measure in our experiment is the Bayesian slope per condition, i.e. the weight subjects place on current sensory information (likelihood) above prior beliefs (where *1* indicates full reliance on likelihood information and none on prior, and *0* indicates the opposite; see Methods for details). Thus, we would expect a smaller difference in the Bayesian slopes between the large prior uncertainty (P) and the small prior uncertainty (p) conditions. We do not find evidence for a different sensitivity to prior uncertainty between PD patients (*off-state*) and controls ($p=0.86$, t-test, n=15; see Figure 3.3), or between *off* and *on* states ($p=0.23$, paired t-test, n=15). PD patients seem to still be able to learn the different prior uncertainties as well as age-matched controls.

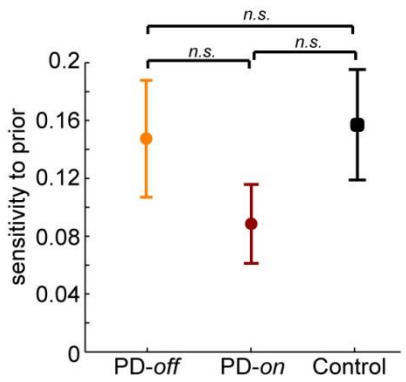


**Figure 3.3 - Sensitivity to prior uncertainty.** It is calculated as the average difference in the Bayesian slope between the 2 types of prior uncertainty. No significant differences were found (p>0.05 for all comparisons). Error bars are SEM.

**No significant difference in average weight given to current vs. prior information**

We also wanted to know if PD affects the average weight given to current sensory information versus prior beliefs, i.e. the average Bayesian slope. We did not find a significant difference in terms of average slope between PD patients (*off-state*) and controls ($p=0.47$, t-test, n=15; see Figure 3.4-A), or between PD patients in *on* vs. *off-state* ($p=0.35$, paired t-test, n=15; see Figure 3.4-B). Analyzing PD patients (*off-state*) and controls, we found a significant main effect of both prior and likelihood type ($F_{(1,29)}>29$, $p<0.001$; see Figure 3.4-B), with subjects placing a higher weight on likelihood when likelihood uncertainty decreased or prior uncertainty increased. Qualitatively, this behavior would be expected if subjects are using a Bayesian strategy (Vilares and Kording, 2011). There was no significant interaction between population and prior types ($F_{(1,29)}=0.03$, $p=0.86$), which agrees with the result shown previously that PD patients can distinguish differences in prior uncertainty as well as age-matched controls. However, there was a significant interaction effect between population and likelihood type ($F_{(1,29)}=11.6$, $p=0.002$), i.e. patients and controls reacted differently to changes in likelihood uncertainty. Our results did not find an effect of PD on the average weight given to current vs. prior information, but show that PD patients and controls react differently to changes in current sensory uncertainty.

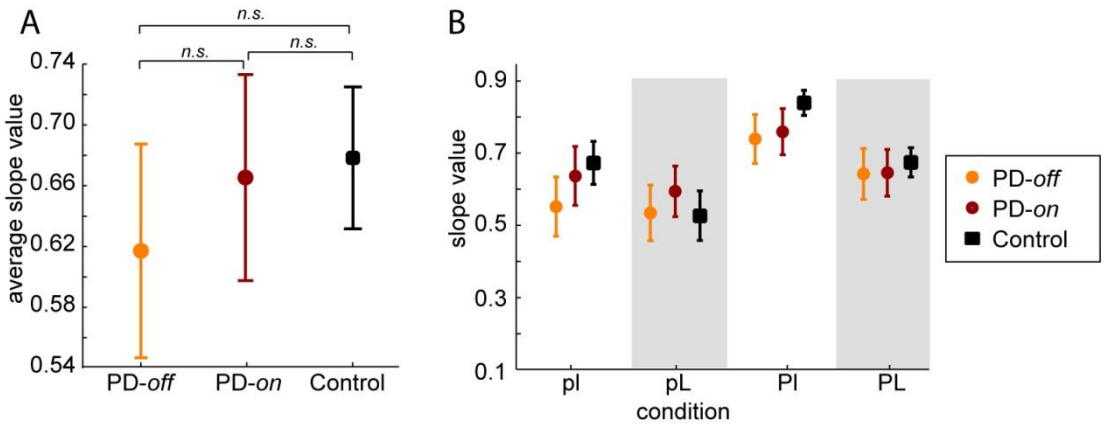


**Figure 3.4 - Weight given to current vs. prior information (Bayesian slope), by condition.** The slope quantifies the degree to which subjects rely on the current visual stimulus (likelihood) versus the prior, with 1 corresponding to full reliance on prior and none on likelihood, and 0 corresponding to the opposite. Average Bayesian slope ± SEM for PD patients *off-state* (orange circles), PD patients *on-state* (red circles) and controls (black squares), averaged across conditions (A) or separated by condition (B). No significant differences were found for the average Bayesian slope ($p>0.05$ for all comparisons).

### PD patients are less sensitive to current sensory uncertainty

To understand how PD patients react to changes in likelihood uncertainty (uncertainty in the current stimulus) we compared the slope between conditions of different likelihood uncertainty. We found that patients (*off-state*) reacted significantly less to changes in likelihood uncertainty compared to controls ($p=0.002$, t-test, n=15; see Figure 3.5). When the prior was narrow (more precise), the difference in behavior was salient and the change in likelihood uncertainty had no significant influence for patients ($p=0.55$, paired t-test, n=15; see also Figure 3.4-B), while controls showed a solid effect ($p=0.001$, paired t-test, n=15). When the prior was wide (more uncertain), both groups showed an effect of likelihood uncertainty ($p=0.002$ and $p<0.001$ for patients and controls, respectively, paired t-test, n=15).

Patients *on-state* did not differ significantly from either *off-state* (*p=0.53*, paired t-test, n=15; see Figure 3.5) or controls (*p=0.052*, t-test, n=15; see Figure 3.5) in their sensitivity to likelihood uncertainty. Together, our results indicate that PD patients, especially *off-state*, are less responsive to changes in the uncertainty of the current stimulus.

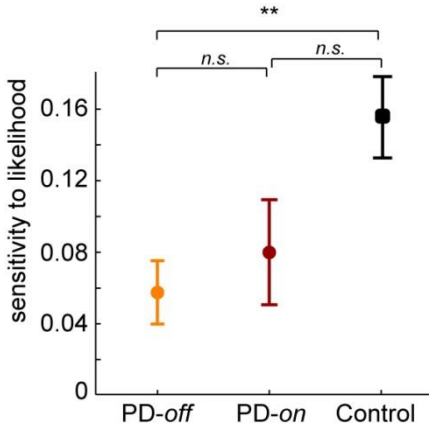


**Figure 3.5 - Sensitivity to likelihood uncertainty.** It is calculated as the average difference in the Bayesian slope between the 2 types of likelihood uncertainty. PD patients o*ff-state*, relative to controls, are significantly less sensitive to differences in likelihood uncertainty (*p=0.002*, t-test, n=15). Error bars are SEM.

### 3.5 Discussion

In this paper, we wanted to understand how Parkinson's disease affects decision-making under uncertainty. We found that PD patients performed worse than age-matched controls in a task in which both prior and visual likelihood uncertainty was present, and that performance improved with administration of dopaminergic drugs. Looking at the way in which they detect and process uncertainty, we found that patients could learn prior uncertainty as well as controls and that they did not differ significantly in the average weight given to current vs. prior information. However, they showed diminished sensitivity to likelihood uncertainty, which was particularly pronounced in the context of a more certain prior. Parkinson disease specifically affected patients' capacity to react to changes in current sensory uncertainty.

If dopaminergic activity from the putamen directly encoded prior uncertainty, then we would have expected PD patients to be impaired at learning the prior uncertainty. However, while PD patients *off-state* showed impairments in performance, they were still able to learn the prior uncertainty as well as controls. Dopamine and the putamen have been previously implicated in learning of some tasks (Dayan and Balleine, 2002, McClure et al., 2003, Frank et al., 2004). Notably, however, they do not seem to be required for learning of all types of tasks. For example, in rodents, inhibiting neural activity from the dorsal striatum, which comprises both putamen and caudate, decreased performance in a task but not learning (Atallah et al., 2007). Similarly, rats in which the tonic dopamine levels were altered could still learn (Beeler et al., 2010). The effects of dopaminergic medication on learning also seem to depend on the specific task (Cools and Robbins, 2004, Frank, 2005). For example, in a study with PD patients, it was also found that dopaminergic drug state (*on vs. off*) impacted performance but not learning (Shiner et al., 2012). Dopaminergic medication increased learning in some tasks but actually decreased learning

in others (Frank et al., 2004, Frank et al., 2007, Moustafa et al., 2008, Pizzagalli et al., 2008). Thus, while dopaminergic activity from the putamen is important for some types of learning, our results indicate that it is not crucial for learning the prior uncertainty, suggesting that it does not directly encode prior uncertainty.

If, as it has been proposed, dopamine activity affected the general weight given to current bottom-up sensory evidence (likelihood) vs. top-down priors, i.e. the Bayesian slope (Beeler et al., 2010, Friston et al., 2012, Galea et al., 2012), then we would expect a difference between PD patients and controls in their average Bayesian slope. We did not find support for this hypothesis. This may be due to a lack of statistical power given the heterogeneity of our sample, which included different stages of the disease, different medicine, and different ages; all factors that produce relevant heterogeneity in basal dopamine levels and hence on its effects (Cools and Robbins, 2004, Chowdhury et al., 2012). It may also be due to a ceiling effect, given that many subjects exhibited a high slope in some of the conditions. Nevertheless, the effects on the sensitivity to differences in likelihood uncertainty were much higher than the potential effects on the average Bayesian slope. Higher reliance on current stimuli and increased attention to it are probably factors that are often correlated in nature. However, it is worth noticing that increased attention to an uncertain stimulus may actually make someone rely less on it than if one was not as attentive and hence did not notice that the stimulus was uncertain. Our results do not support the role of dopamine in signaling the general weight given to current vs. prior information.

If dopamine signals the level of current sensory uncertainty, putatively through increases of salience or attention given to the sensory stimulus, we hypothesized that PD patients should be impaired at reacting to differences in its uncertainty. Our results strongly supported this hypothesis. These results agree with a series of experiment and theories

that have been proposed about dopamine (Berridge and Robinson, 1998, Kapur, 2003, McClure et al., 2003, Dang et al., 2012). In this view, a given sensory stimulus, either because of its novelty, intensity, or potential importance for an animal's fitness, will be particularly salient and activate a dopaminergic response that leads the animal pays more attention to it (Redgrave and Gurney, 2006). We propose that this increased attention will also enable the animal to better detect the uncertainty in the sensory stimulus, and hence respond appropriately to it. This is in accordance with the interpretation that attention is the inference about the uncertainty (precision) of a sensory signal (Feldman and Friston, 2010). It is also in accordance with a study that found that there is a specific loss of attentional mechanisms in PD which is independent of visual sensory deficits (Sampaio et al., 2011). Our data supports the role of dopamine in regulating attention to sensory stimuli as a means of signaling their uncertainty.

The lack of sensitivity to the uncertainty in the current stimulus was particularly pronounced in the context of a more certain prior. A more uncertain prior may in itself also be a drive for increased attention to current stimuli, and hence could somewhat compensate for the patients' general lack of attention to it. This also agrees with the results obtained in Galea *et al.* (2012), which showed that PD patients were slower to change behavior in response to an improbable stimulus when this stimulus was delivered in the context of an overall predictable sequence (i.e. a more certain prior), but not when the sequence was unpredictable (uncertain prior) (Galea et al., 2012). It is of course possible that the dopaminergic activity from the putamen directly codes for the uncertainty (or its inverse, precision) in the current stimulus (Friston et al., 2012). However, if that would be the case then we would again expect a generalized increased/decreased reliance on likelihood. Previous research also does not support this hypothesis (McKeefry et al., 1997, Vilares et al., 2012). The observed asymmetric sensitivity to likelihood uncertainty depending on the characteristics of the

prior further reinforces the idea that dopamine may not directly code for uncertainty in the current stimulus but, instead, help in signaling its uncertainty by increasing attention to it.

We cannot be sure that the cause for the impairments we observed in PD patients is specifically related to the putamen, as transmission from other structures is also affected. Nevertheless, the putamen seems to be the brain structure most severely compromised in Parkinson's disease (Kish et al., 1988), and so it is likely that the observed effects are at least partly due to a deficient dopaminergic transmission from the putamen. It would be interesting to perform a similar study with other patient populations that also have deficits in transmission from the putamen, such as Wilson's disease (King et al., 1996). We predict that if the observed impairments are specifically related to activity from the putamen, then these patients will also show deficits in detecting differences in likelihood uncertainty.

We did not find a significant improvement in the sensitivity to likelihood uncertainty after dopaminergic replacement therapy administration (i.e. when subjects where "on state"). Again this may be related with the high variability in our sample in terms of age, stage of the disease and type of dopaminergic medication, causing potentially more variable results (Cools and Robbins, 2004). Also, the effect of dopamine on a particular brain system follows an "inverted U-shaped" function, with both too much and too little dopamine impairing behavior, and the concentrations of dopamine given to the patients may not have been the optimal ones. Finally, dopaminergic drugs are administered outside the cells and probably do not completely replace the internally generated dopamine effects. It would be interesting to do a similar task with either never medicated patients or healthy controls, and see the effect different levels of dopamine administration could have on behavior.

The computational role of dopamine proposed here may help in the understanding of some of the symptoms both PD and schizophrenic patients show. It has been proposed that some of the schizophrenia symptoms can be related to an incorrect Bayesian inference process (Fletcher and Frith, 2009). We subscribe this view, and suggest that they may be associated with the role of dopamine in attending to the current sensory stimulus and its uncertainty. Schizophrenics are suggested to have a deregulated dopamine transmission that leads to stimulus-independent release of dopamine (Kapur, 2003). Hence, this would lead the person to pay more attention to current sensory stimuli, even when there is no apparent reason for it. Some reports made by schizophrenics suggest exactly that: “Sights and sounds possessed a keenness that he had never experienced before”; or “My senses seemed alive…. I noticed things I had never noticed before (Bowers and Freedman, 1966, Bowers, 1968, Kapur, 2003). On the other side of the spectrum we have never medicated PD patients, who do not seem to have any interest in novel things (Bodi et al., 2009). If, as existing literature and our paper suggest, dopamine promotes paying attention to current sensory stimuli, this may help in the awareness of the specific difficulties both PD and schizophrenic patients face. In a way, we hope that our paper serves the role of dopamine itself.

The neural signaling of uncertainty is among the central topics in computational neuroscience (Ma et al., 2006, Feldman and Friston, 2010, Fiser et al., 2010, Vilares and Kording, 2011). However, preciously little is known about the way the brain actually solves those problems. There exists a growing literature using fMRI studies asking how neural activities change with uncertainty (Critchley et al., 2001, Hsu et al., 2005, Behrens et al., 2007, Vilares et al., 2012, O'Reilly et al., 2013). However, these studies cannot directly address causal roles of brain areas or neurotransmitters. Here we have used a different approach and started with a patient population that should, according to multiple theories, have deficits in the

signaling of uncertainty. Our findings of retained learning, unbiased slopes, but a lack of sensitivity to changes in current sensory uncertainty, promise to inform the development of new theories about the representation of uncertainty.

### 3.6 Acknowledgements

We would like to thank David Klein, Laura Pickering, Dr. Santiago Toledo, Dr. Citlali López-Ortiz and especially Dr. Tania Simuni for help in recruiting Parkinson's disease patients. We also would like to thank Peter Dayan and Hugo Fernandes for useful comments on the manuscript.

### 3.7 References

Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC (2007) Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. Nature Neuroscience 10:126-131.

Beeler JA, Daw N, Frazier CRM, Zhuang XX (2010) Tonic dopamine modulates exploitation of reward learning. Front Behav Neurosci 4.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nature Neuroscience 10:1214-1221.

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Rev 28:309-369.

Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibo G, Takats A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the

effects of dopamine agonists on young Parkinsons patients. Brain 132:2385-2395.

Bowers MB (1968) Pathogenesis of Acute Schizophrenic Psychosis - an Experiential Approach. Arch Gen Psychiat 19:348-&.

Bowers MB, Freedman DX (1966) Psychedelic Experiences in Acute Psychoses. Arch Gen Psychiat 15:240-&.

Brand M, Labudda K, Kalbe E, Hilker R, Emmans D, Fuchs G, Kessler J, Markowitsch HJ (2004) Decision-making impairments in patients with Parkinson's disease. Behav Neurol 15:77-85.

Chowdhury R, Guitart-Masip M, Bunzeck N, Dolan RJ, Duzel E (2012) Dopamine Modulates Episodic Memory Persistence in Old Age. J Neurosci 32:14193-14204.

Cools R, Barker RA, Sahakian BJ, Robbins TW (2003) L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. Neuropsychologia 41:1431-1441.

Cools R, Robbins TW (2004) Chemistry of the adaptive mind. Philos T Roy Soc A 362:2871-2888.

Critchley HD, Mathias CJ, Dolan RJ (2001) Neural activity in the human brain relating to uncertainty and arousal during anticipation. Neuron 29:537-545.

Dang LC, O'Neil JP, Jagust WJ (2012) Dopamine Supports Coupling of Attention-Related Networks. J Neurosci 32:9582-9587.

Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. Neuron 36:285-298.

Delazer M, Sinz H, Zamarian L, Stockner H, Seppi K, Wenning GK, Benke T, Poewe W (2009) Decision making under risk and under ambiguity in Parkinson's disease. Neuropsychologia 47:1901-1908.

Doyon J, Bellec P, Amsel R, Penhune V, Monchi O, Carrier J, Lehericy S, Benali H (2009) Contributions of the basal ganglia and functionally related brain structures to motor learning. Behavioural Brain Research 199:61-75.

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.

Fiser J, Berkes P, Orban G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14:119-130.

Fletcher PC, Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 10:48-58.

Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cognitive Neurosci 17:51-72.

Frank MJ, Samanta J, Moustafa AA, Sherman SJ (2007) Hold your horses: Impulsivity, deep brain stimulation, and medication in parkinsonism. Science 318:1309-1312.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. Science 306:1940-1943.

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293-301.

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S (2012) Dopamine, Affordance and Active Inference. Plos Comput Biol 8.

Galea JM, Bestmann S, Beigi M, Jahanshahi M, Rothwell JC (2012) Action Reprogramming in Parkinson's Disease: Response to Prediction Error Is Modulated by Levels of Dopamine. J Neurosci 32:542-550.

Grafton ST, Hazeltine E, Ivry R (1995) Functional Mapping of Sequence Learning in Normal Humans. J Cognitive Neurosci 7:497-510.

Horvitz JC (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. Neuroscience 96:651-656.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Kapur S (2003) Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiat 160:13-23.

King AD, Walshe JM, Kendall BE, Chinn RJ, Paley MN, Wilkinson ID, Halligan S, Hall-Craggs MA (1996) Cranial MR imaging in Wilson's disease. AJR Am J Roentgenol 167:1579-1584.

Kish SJ, Shannak K, Hornykiewicz O (1988) Uneven Pattern of Dopamine Loss in the Striatum of Patients with Idiopathic Parkinson's Disease. New England Journal of Medicine 318:876-880.

Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244-247.

Lehericy S, Benali H, Van de Moortele PF, Pelegrini-Issac M, Waechter T, Ugurbil K, Doyon J (2005) Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. P Natl Acad Sci USA 102:12566-12571.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432-1438.

McClure SM, Daw ND, Montague PR (2003) A computational substrate for incentive salience. Trends Neurosci 26:423-428.

McKeefry DJ, Watson JDG, Frackowiak RSJ, Fong K, Zeki S (1997) The activity in human areas V1/V2, V3, and V5 during the perception of coherent and incoherent motion. Neuroimage 5:1-12.

Mimura M, Oeda R, Kawamura M (2006) Impaired decision-making in Parkinson's disease. Parkinsonism Relat D 12:169-175.

Moustafa AA, Sherman SJ, Frank MJ (2008) A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. Neuropsychologia 46:3144-3156.

O'Reilly JX, Jbabdi S, Rushworth MF, Behrens TE (2013) Brain Systems for Probabilistic and Dynamic Prediction: Computational Specificity and Integration. PLoS Biol 11:e1001662.

Orban P, Peigneux P, Lungu O, Albouy G, Breton E, Laberenne F, Benali H, Maquet P, Doyon J (2010) The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. Neuroimage 49:694-702.

Pizzagalli DA, Evins AE, Schetter EC, Frank MJ, Pajtas PE, Santesso DL, Culhane M (2008) Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. Psychopharmacology 196:221-232.

Playford ED, Jenkins IH, Passingham RE, Nutt J, Frackowiak RSJ, Brooks DJ (1992) Impaired Mesial Frontal and Putamen Activation in Parkinsons-Disease - a Positron Emission Tomography Study. Ann Neurol 32:151-161.

Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, Knowlton BJ (2005) The neural correlates of motor skill automaticity. J Neurosci 25:5356-5364.

Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci 7:967-975.

Sampaio J, Bobrowicz-Campos E, André R, Almeida I, Faria P, Januário C, Freire A, Castelo-Branco M (2011) Specific impairment of visual spatial covert attention mechanisms in Parkinson's disease. Neuropsychologia 49:34-42.

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, Dolan RJ (2012) Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. Brain 135:1871-1883.

Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE (2010) Systematic Review of Levodopa Dose Equivalency Reporting in Parkinson's Disease. Movement Disord 25:2649-2653.

Vilares I, Howard James D, Fernandes Hugo L, Gottfried Jay A, Kording Konrad P (2012) Differential Representations of Prior and Likelihood Uncertainty in the Human Brain. Current Biology 22:1641-1648.

Vilares I, Kording K (2011) Bayesian models: the structure of the world, uncertainty, behavior, and the brain. Year in Cognitive Neuroscience 1224:22-39.

Wise RA (2004) Dopamine, learning and motivation. Nat Rev Neurosci 5:483-494.

# 4. TRUST GAMES: ARE MONEY AND EFFORT EQUIVALENT?

# 4. TRUST GAMES: ARE MONEY AND EFFORT EQUIVALENT?

**Citation:** Vilares I, Dam G and Kording KP (2011). Trust games: Are money and effort equivalent? PLos ONE, 6(2): e17113.

**Author Contributions:** All authors contributed in designing the study. IV and GD performed the experiments; IV analyzed the data; IV and KK wrote the paper with input from GD.

## 4.1 Summary

Trust and reciprocity facilitate cooperation and are relevant to virtually all human interactions. They are typically studied using trust games: one subject gives (entrusts) money to another subject, which may return some of the proceeds (reciprocate). Currently, however, it is unclear whether trust and reciprocity in monetary transactions are similar in other settings, such as physical effort. Trust and reciprocity of physical effort are important as many everyday decisions imply an exchange of physical effort, and such exchange is central to labor relations. Here we studied a trust game based on physical effort and compared the results with those of a computationally equivalent monetary trust game. We found no significant difference between effort and money conditions in both the amount trusted and the quantity reciprocated. Moreover, there is a high positive correlation in subjects' behavior across conditions. This suggests that trust and reciprocity may be character traits: subjects that are trustful /trustworthy in monetary settings behave similarly during exchanges of physical effort. Our results validate the use of trust games to study exchanges in physical effort and to characterize inter-subject differences in trust and reciprocity, and also suggest a new behavioral paradigm to study these differences.

*"But if you could not trust, you were indeed alone, for neither friendship nor partnership, neither family nor alliance could exist without it. Without trust, we were scattered far and wide, at the mercy of the four winds with nothing to cling to."*

*Julliet Marillier*

## 4.2 Introduction

Trust exists to some degree in all human interaction and is associated with a more healthy, egalitarian and productive society (Knack and Keefer, 1997, Camerer, 2003, Krueger et al., 2007). It can be defined as a positive expectation in the face of uncertainty emerging from social relations (Gambetta, 2000, Guseva and Rona-Tas, 2001). Trust enables cooperative behavior, facilitates organization in both permanent and temporary work groups and is associated with higher job satisfaction, lower labor cost and larger profits (Meyerson et al., 1996, Knack and Keefer, 1997, Gambetta, 2000, Chami and Fullenkamp, 2001). It can be seen in diverse types of interactions: people trust money to bankers, hoping that they won't run away with it; they trust their own home, letting complete strangers stay in their house and they also trust physical effort, for example by helping a friend move (Guseva and Rona-Tas, 2001, Lauterbach et al., 2009). Across such situations trust plays different roles and it seems important to understand commonalities and differences.

Trust is often justified as humans express reciprocity: they return helpful or harmful acts in kind, even though such behavior may come at a cost (Camerer and Fehr, 2002, Nowak and Sigmund, 2005). As in trust, reciprocity is also expressed in different situations: in the examples above, the banker will work harder to maximize the earnings of the trustful investor,

and a person that just stayed for free in someone else's house will more likely consider hosting as well (Lauterbach et al., 2009). If two people interact repeatedly, then not reciprocating but exploiting the partner has to be weighted against the cost of losing collaboration in the future (McCabe et al., 2001, Nowak and Sigmund, 2005). If, however, partners only interact once, then there is no direct risk of such retaliation. Nevertheless, even in one-shot interactions humans tend to reciprocate, while this behavior is much more difficult to find in other species (Nowak and Sigmund, 2005). Reciprocity in single encounters is of special interest for economists since in the current global market the traditional long-term repeated interactions between relatives or neighbors are being slowly replaced by one-time interactions between anonymous partners (Nowak and Sigmund, 2005). A better understanding of reciprocity in one-shot interactions in all its different contexts can then be of particular relevance to the current economy.

The importance of trust and reciprocity has been progressively recognized in the field of labor economics. Trust increases the ability of group members to work together (Dirks, 1999) and promotes reciprocity (Fehr and List, 2004). It also seems to affect effort. For example, intensive control by a supervisor may lead to decreased work effort because it is sensed as an indication of distrust (Frey, 1993, Guerra, 2002). Furthermore, trust within a group seems to affect their work effort, although the relationship between trust and effort is not very clear (Dirks and Ferrin, 2001). Some studies suggest that higher levels of trust can increase effort and efficiency towards the group task (Klimoski and Karol, 1976, Dirks, 1999). Paradoxically, it has also been proposed that in some situations people with low trust will actually work harder when in a group, in order to compensate for the putative low performance of the co-workers (Williams and Karau, 1991). Reciprocity, by its turn, not only reinforces trust but it can also increase employees working effort, functioning as an effective contract-enforcement device (Fehr et al., 1997, Gachter and Falk, 2002). These

studies thus indicate that trust and reciprocity can affect effort, but how do they relate is still unclear.

In the context of behavioral economics, trust and reciprocity are often studied using trust games (Berg et al., 1995, Camerer, 2003, Cesarini et al., 2008). In such a game, one individual (the trustor) receives a given amount of money, and can choose how much of it to trust or invest. The trusted amount is then multiplied by some factor, for example three (symbolizing a return on social investment), and given to the other player (the *trustee*). The trustee can then decide how much of the proceeds to keep and how much to return to the trustor. The amount of money invested by the trustor is a measure of trust, and the amount repaid back by the trustee is a measure of reciprocity. In this way, trust games allow quantifying both the degree of trust as well as the degree of reciprocity.

Several studies have used trust games and the results have been contrary to what would be expected under the assumption of purely self-interested individuals, who act in order to maximize their own payoff (Camerer and Fehr, 2002). In fact, if the trust game is played only once, then the optimal strategy of a purely self-interested trustee is to not reciprocate any money, and so the trustor, anticipating this, would invest nothing (Camerer, 2003, King-Casas et al., 2005). Thus, for a one-shot trust game, the Nash equilibrium (the solution in which no player can increase their payoff unilaterally) is to neither trust nor reciprocate. Instead, it has been found that people do trust and reciprocate even at a cost to their gains (Berg et al., 1995, Fehr and Gachter, 2000, Camerer, 2003). However, these results were generally obtained using exchanges of money, and how this “monetary trust and reciprocity” can extrapolate to other contexts, such as effort, is still largely unknown (Levitt and List, 2007).

Trust and reciprocity of effort may have different properties when compared to trust and reciprocity of money. Studies have suggested that

trading effort instead of money can lead to different results, as it might increase the amount of cooperation (van Dijk et al., 2001) and can affect property rights (Fahr and Irlenbusch, 2000). Even in daily life this might be seen. For example, many people would easily give a day's worth of work to help a friend to move, but would not so easily offer them an equivalent amount of money (van Dijk et al., 2001). There are thus indications that people are willing to entrust effort more than money. If trust and reciprocity differ between exchanges of money and effort, then caution is necessary when generalizing the results of the monetary trust games to the domain of effort.

In this study, we wanted to know how people trust and reciprocate effort in the context of a trust game. Specifically, we focused on physical effort, as it can be readily measured. Each of our 60 subjects participated in two computationally equivalent trust games, one involving physical effort and one involving money. We found that there were no significant differences in trust and reciprocity between effort and monetary conditions. These results hold even if we analyze only the first game of each subject. We also found that, across the two conditions, trust and reciprocity were strongly correlated. Finally, we observed that, although on average subjects reciprocated identically in both effort and monetary conditions, there was a much higher variability in the proportion reciprocated for the effort condition.

## 4.3 Materials and methods

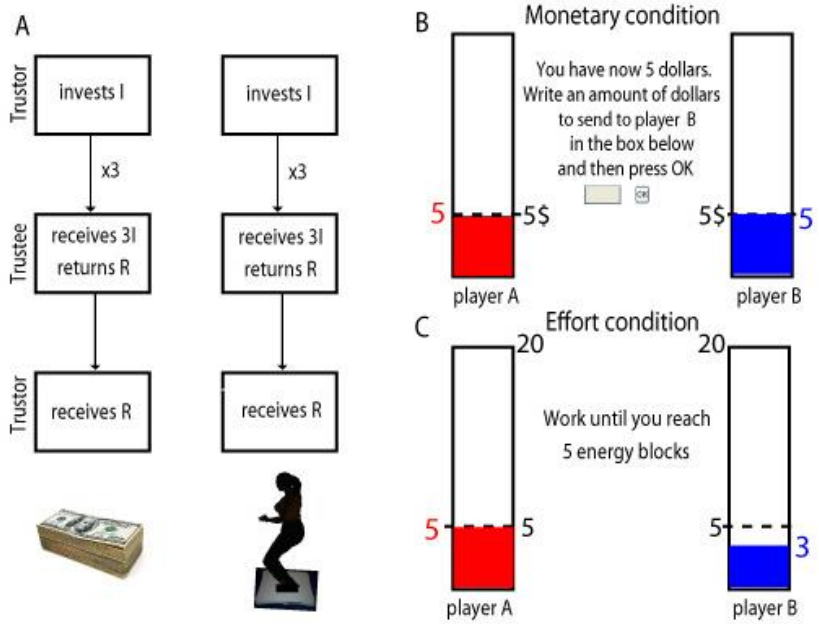


**Figure 4.1 - Experimental setup.** A) Representation of the trust game used in this experiment. In each condition the trustor sends some amount I between 0 and 5 units, units which in the monetary condition (on the left) are US $ and in the effort condition (on the right) are energy blocks (EB). This amount is then tripled and sent to the trustee, who can then return some quantity R of this tripled amount (3I). In the effort condition both players have then to do the remaining squats in order to arrive to 20 EB. Each subject plays each condition only once. They keep the same role throughout the entire experiment but change partners between conditions. B) Example image presented on the computer screen when subjects were performing the monetary condition part of the experiment. In this phase, the trustor ("player A") had just received the 5 $ show-up fee and had to decide how much to send to the trustee ("player B"). C) Example image presented on the computer screen when subjects were performing the effort condition part of the experiment. In this phase, the trustee ("player B") was performing squats in order to arrive to 5 EB. While squatting, the correspondent bar (blue for the trustee, red for the trustor) was going up and the total number of EB possessed by the trustee was shown (as a blue number next to the trustee's rectangle). Once the trustee reached the 5 EB threshold the amount given by the trustor would be shown.

***Participants:*** A total of 60 healthy volunteers (35 females, 25 males; age 30 ±9 years) participated in the experiment. All experimental protocols were performed in accordance with federal guidelines and the Northwestern University's policy statement on the use of humans in experiments. Informed consent was obtained from all participants.

***Experimental procedure:*** Each experimental session consisted of a trust game presented in 2 conditions, monetary and physical effort. All subjects in the same session experienced the same sequence of events. The choice to start with the effort or the monetary condition was pseudorandom, so that half of the volunteers started with one condition, and the other half started with the other, to avoid any potential priming effects. Each subject played both conditions of the trust game once. Subjects were randomly assigned to be either player A (trustor) or player B (trustee) and kept that role throughout the experimental session. In order to minimize potential multi-round effects, subjects were randomly matched for the first condition of the experiment and changed partners between conditions. Players with the same role were placed in the same room, and were informed that they would only be playing with subjects that were in another room. They were instructed to not discuss strategies with one another. Furthermore, they were also informed that the person with whom they were playing would change between experiments. Subjects were given no information that enabled them to identify their partner and were also asked to keep their decisions private. They received written instructions. The instructions and the computer screen were phrased as neutral as possible; words like "trust", "cooperation", "competition" and "opponent" were avoided. An experimental session (including instructions, both experiments and waiting time) averaged 60 minutes. Earnings averaged $18 and ranged between $14 and $25.

***The Game:*** In the *monetary condition*, both player A (trustor) and player B (trustee) received 5 US dollars ($5) as a show-up fee. Player A then decided to send all, none or some (in multiples of $0.25) of the show-up fee to player B. The amount sent to player B was tripled. Player B then decided how much of that money to send back to player A and how much to keep (see Figure 4.1-A). In the *effort condition*, both player A and player B had to perform squats while standing on a 4-sensor force-plate until each of them reached 20 “energy blocks” (EB). Initially both player A and player B had to perform squats until a total of 5 energy blocks was reached. Player A then had the opportunity of sending a portion of the 5 energy blocks (in multiples of 0.25) to player B. The amount of energy blocks sent to player B was tripled. Player B then decided how many energy blocks to return to player A (see Figure 4.1-A). Each player had then to perform the remaining squats in order to reach the 20 energy blocks required for the task. Each player received $10 for completing this part of the experiment.

***Data acquisition:*** Participants wrote their decisions in a box on a computer screen. At each point of time, the computer screen showed two rectangles, one at the left side of the screen with the amount of dollars/energy blocks possessed by player A (trustor) and one at the right side with the amount of dollars/energy blocks possessed by player B (trustee). Furthermore, a color-code was given: red for the dollars/energy blocks earned by the trustor and blue for the ones earned by the trustee (see Figure 4.1-B and Figure 4.1-C). Subjects’ responses were recorded using Matlab. For the effort experiment, data was collected using a 4-sensor force-plate (Nintendo Wii™ Balance Board, recorded at 500Hz). By performing squatting movements with their body, subjects produced forces, which were then translated to Energy Blocks.

An energy block (EB) is a multiple of the work produced by that subject (W) per unit of mass (m):

$$EB = c\frac{W_t}{m}$$

Where *c* is a constant, in this case *c=1/4*. Work was defined at each point of time *t* based on the forces recorded via:

$$W_t = \sum_{i=0}^{t} |F_i|\,|v_i|\,\Delta t_i$$

Where $F_i$ is the average force obtained by the 4-sensor force plate (in *Newtons*) at each iteration *i,* $v_i$ is the velocity and $\Delta t_i$ is the amount of time that passed between *i-1* and *i* (which was, on average, about 0.035 s).

### 4.4 Results

In this study we asked if trust and reciprocity differ between equivalent monetary and effort conditions of the trust game (see Figure 4.1). In one condition subjects traded money (in units of US$) and in the other they traded effort, which was measured in energy blocks (*EB*). We considered the amount sent by the trustor as absolute trust and the amount sent back by the trustee as absolute reciprocity, or trustworthiness. We also considered relative trust, the amount sent by the trustor divided by the total amount available (in our case, the total amount the trustor had available was 5 $ or 5 *EB*) and relative reciprocity, the amount returned by the trustee divided by the total amount available (i.e. three times the amount sent by the trustor). Every subject participated in each condition once, changing partners between conditions. Subjects kept their role (trustor or trustee) throughout the experiment so that we could compare how they trusted or reciprocated across conditions.

**No significant difference in trust between monetary and effort conditions**

We sought to test whether trust differs between monetary and effort conditions. We found that subjects displayed trusting behavior in both conditions (Figure 4.2-A), trusting $\mu_{\$}$=3.1±0.3 $ in the monetary condition and similarly $\mu_{w}$=3.1±0.3 *EB* in the effort condition. Both averages were significantly different from zero, which is the Nash equilibrium for this game (*p-val*$_{\$}$=1.6x10$^{-6}$ and *p-val*$_{w}$ =1.5x10$^{-6}$, Wilcoxon signed-rank test). No significant difference in trust was found when comparing the monetary and effort conditions (*p-val*= 0.96; Wilcoxon signed-rank test; n=30). Comparing the distribution functions (see Figure 4.2-B) also no difference can be found (*p-val=1*, two-sample Kolmogorov-Smirnov test). This lack of difference between trust in the monetary and effort conditions surprised us, as we had expected subjects to trust more in the effort condition of the task.

We wondered if the lack of a statistical difference was due to an insufficient sample size. We therefore performed a power analysis asking which effect size we should have detected 90% of the time, at a $\alpha$=.05 level of significance. We found that with the variance present in the data and the number of subjects used we should have been able to observe a difference if it had exceeded 15%. Therefore, a difference between both conditions, if existent, should be smaller than this value. Trusting behavior in our experiment, thus, seems very similar between the monetary and the effort conditions.

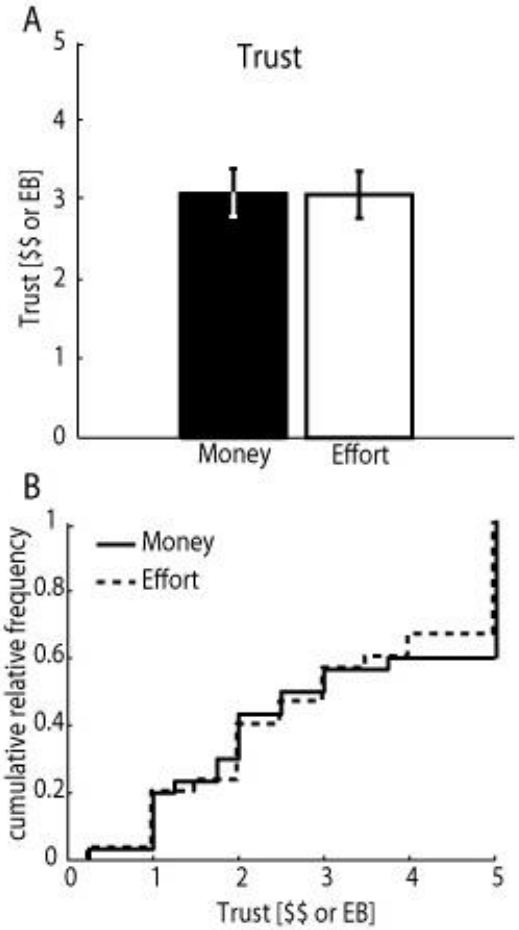


**Figure 4.2 - Quantity sent by the trustor, which is considered a measure of trust. A)** Average trust in the monetary and the effort conditions (black and white, respectively). No significant difference in trust was found comparing the monetary and effort conditions (*p-val*= 0.9587; Wilcoxon signed-rank test; n=30). Error bars represent standard error of the mean (s.e.m.). **B)** Cumulative distribution function for the amount sent by the trustor for the monetary (solid line) and the effort (dashed line) conditions. The distribution functions are not statistically different (*p=1*, two-sample Kolmogorov-Smirnov test; n=30).

**No significant difference in reciprocity between monetary and effort conditions**

We also wanted to know if reciprocating behavior differs between a monetary and an effort condition. We found that subjects reciprocated in both conditions (Figure 4.3-A), returning $\mu_{\$}$=4.9 ±0.6 $ in the monetary condition and $\mu_w$=4.8±0.6 *EB* in the effort condition. The averages were again significantly different from zero, the Nash equilibrium (p-$val_{\$}$=1.6x10$^{-6}$ and p-$val_w$ =2.5x10$^{-6}$, Wilcoxon signed-rank test). The same is observed if we look at the relative reciprocity ($\mu_{\$}$=0.49 ±0.03$ and $\mu_w$=0. 54±0.05 EB, *p-val*$_{\$}$=1.5x10$^{-6}$ and *p-val*$_w$ =2.4x10$^{-6}$, Wilcoxon signed-rank test). No significant difference in absolute reciprocity was found comparing the monetary and effort conditions (*p-val* = 0.95; paired t-test). Comparing relative reciprocity, there might be a trend for reciprocating more in the effort condition, but the difference is not significant (*p-val* =0.27; paired t-test). Looking at the cumulative distribution functions of monetary versus effort reciprocity (see Figure 4.3-B) no significant difference can be found (*p-val* =0.76, two-sample Kolmogorov-Smirnov test). A power analysis (as above) gives an upper limit of 14%, so any real difference, if it exists, should be smaller than that value. These results indicate that reciprocity does not differ significantly between the monetary and effort conditions, and in both cases subjects reciprocated more than predicted by Nash equilibrium.

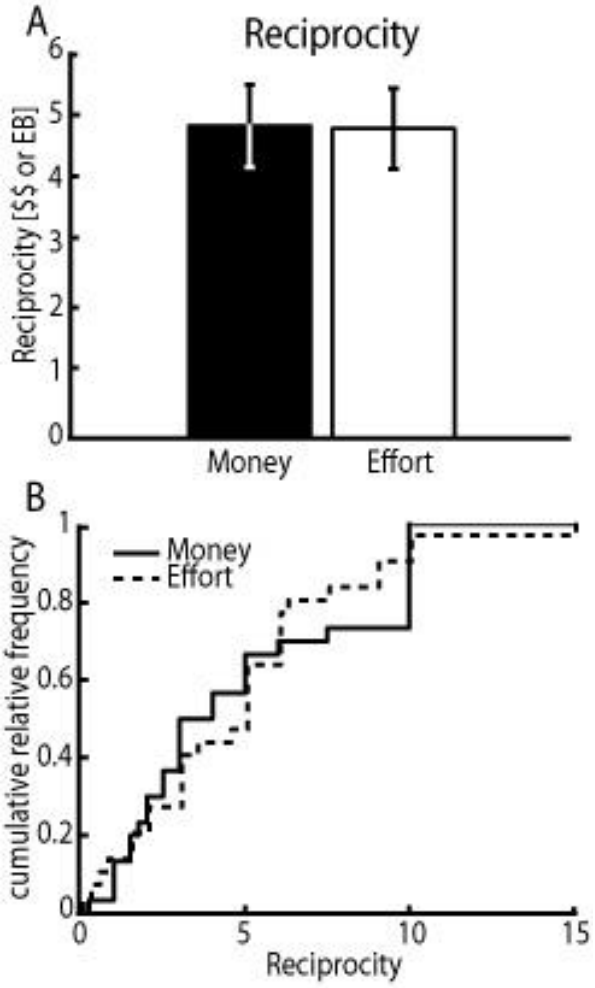


**Figure 4.3 – Quantity sent back by the trustee, which is considered a measure of (absolute) reciprocity.** A) Average reciprocity in the monetary and the effort conditions (black and white, respectively). No significant difference was found comparing the monetary and effort conditions (p= 0.9519; paired t-test; n=30). Error bars represent s.e.m. . B) Cumulative distribution function for the amount returned by the trustee for the monetary (solid line) and the effort (dashed line) conditions. The distribution functions are not statistically different (*p*=0.76, two-sample Kolmogorov-Smirnov test; n=30).

**Significant individual correlations**

We wanted to test if, at an individual level, a subject's behavior in the monetary condition was correlated to behavior in the physical effort condition. We designed the experiment so that each subject participated in one monetary and one effort condition in random order and thus we can analyze correlations across the conditions. We found a high positive correlation in subjects' trusting behavior between the monetary and effort conditions (r=0.74, *p-val* $<10^{-5}$, spearman correlation; see Figure 4.4). We also found a significant but weaker correlation in reciprocating behavior (r=0.39, *p-val* =0.032 for absolute reciprocity; r=0.48, *p-val* =0.008 for relative reciprocity; spearman correlations, n=30; see Figure 4.4). Thus, subjects' behavior was positively correlated between conditions, with subjects that trusted or reciprocated more in a monetary condition tending to be more trusting /trustworthy in a physical effort condition.

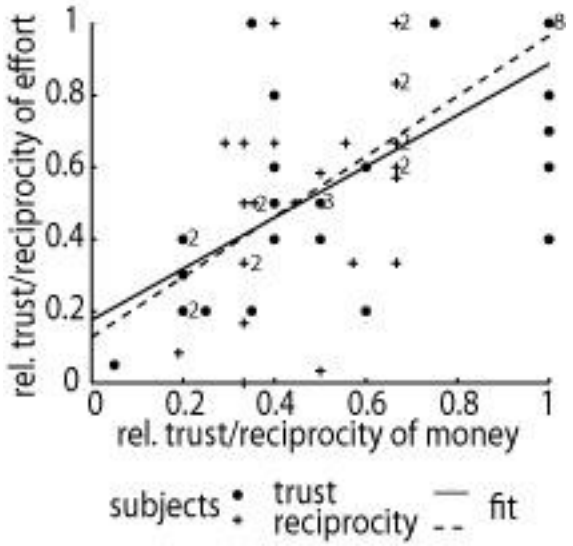


**Figure 4.4 - Relationship between a subject's relative trust or relative reciprocity in the monetary and effort conditions.** Each trustor is represented by a black dot (n=30) and each trustee is represented by a grey cross (n=30). When dots or crosses overlap a small number is shown nearby, representing the amount of overlapping dots (in black) or crosses (in grey). The black solid line represents a linear regression of the relative amount trusted in the effort condition as a function of the relative trust in the monetary condition ($\beta$=0.71, *p-val*=1.6x10$^{-6,}$ $r^2$=0.57). The grey dashed line represents a linear regression of the relative amount reciprocated in the effort condition as a function of the relative amount reciprocated in the monetary condition ($\beta$=0.84, *p-val*=6.9x10$^{-3}$, $r^2$=0.23).

To better understand how people reciprocate trust, we analyzed the correlations between subjects' investment and what they received back from the trustee. We found significant positive correlations between the amount subjects trust and what they receive back from the trustee, for both the monetary ($r_{\$}$=0.93, *p-val*$_{\$}$ <10$^{-8}$) and the effort ($r_{w}$ =0.73, *p-val*$_{w}$ <10$^{-5}$) conditions. The high correlation between trusted and returned amounts can be expected, since more trust provides more money or energy blocks that can be returned. However, looking at the correlations between the relative

values (relative trust with relative reciprocity), they are much weaker in the monetary condition ($r_{\$}$=0.38, $p\text{-}val_{\$}$=0.04), and disappear in the effort condition ($r_w$ =-0.14, $p\text{-}val_w$ =0.45). Thus, subjects give back more if they receive more, although the relative reciprocity does not appear to depend strongly on the relative trust.

We also wanted to know how trustor's behavior in the second round was correlated to behavior of the trustee in the first round. We found a strong positive correlation (r=0.73, *p-val* $<10^{-5}$, n=30, spearman correlation) between the absolute amount received in the first round and what is trusted in the second round. Thus, as it has been observed before (King-Casas et al., 2005), positive interactions in one round are correlated with cooperative behavior in the next.

To understand if this correlation is due to a causal influence of trustee's behavior in round one on trustor's behavior in round two, we constructed a multiple linear regression model in which both trust and reciprocity in round one are used as predictors of trust in round two. According to the model's fit, trust in the first round influences trust in the second (b=0.58, *p-val*=0.01), which would be expected if trusting is a character trait. We found that the behavior of the trustee in round one, on the other hand, had no significant influence (b=0.14, *p-val*=0.22), which would be expected if trustors take into consideration that they are playing with two distinct individuals. It appears that subjects begin the study with a certain level of trust, which is shared between the monetary and the effort conditions, and that they do not significantly update that level based on experience during the first round.

**Behavior using only the first round**

It is possible that the similarity across rounds occurred because subjects simply decided to behave in the second round in the same way they did during the first. To rule out this hypothesis we compared behavior using only the first round of the game. In this way, every subject only contributes one independent data point (see Figure 4.5). Subjects trusted $\mu_{\$}$=2.8 ±0.4 $ and $\mu_{w}$=3.4±0.5 *EB* and reciprocated a total of $\mu_{\$}$=4.3 ±0.8 $ and $\mu_{w}$=4.6±0.8 *EB* (values significantly different from zero, *p-val* =6.1x10$^{-5}$ for all cases, Wilcoxon signed-rank test, n=15). The relative returns follow the same tendency ($\mu_{\$}$=0.47 ±0.04 $ and $\mu_{w}$=0.50±0.06 *EB*; *p-val* =6.1x10$^{-5}$). Comparing the monetary and effort conditions, there was a tendency for higher trust and reciprocity in the effort condition, but this difference is also non-significant (p=0.35 for trust and p=0.68 or p=0.69 for absolute and relative reciprocity; Wilcoxon rank-sum test, n=15). Thus, the observed similarity across conditions does not seem to be a result of the repeated nature of the experiment.

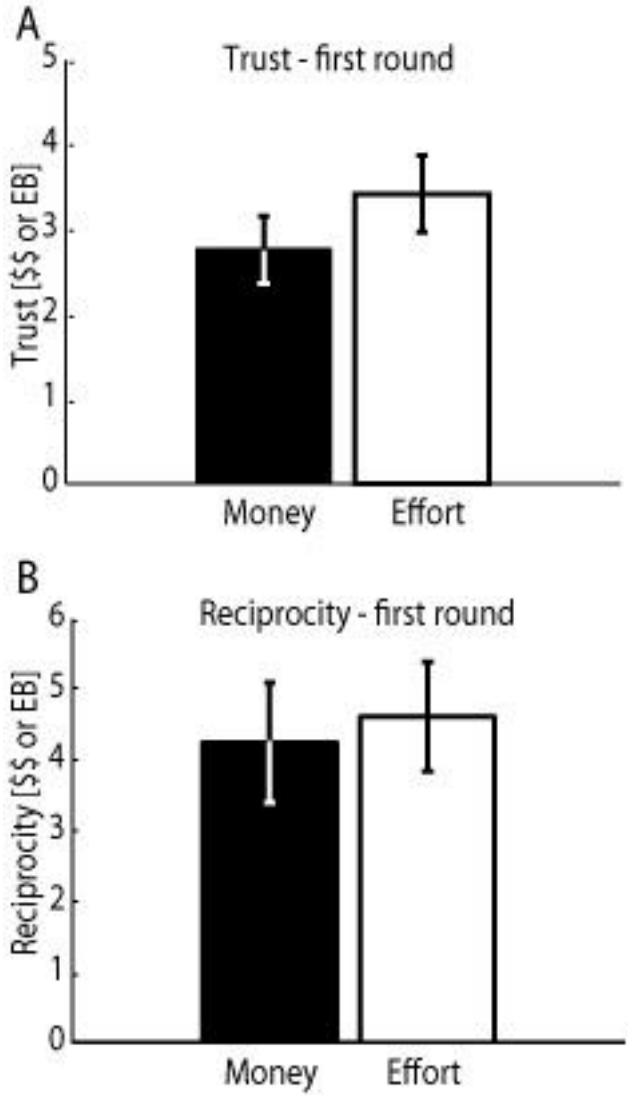


**Figure 4.5 - Trust and reciprocity values using only data from the first round.** A) Average amount sent by the trustor (trust) in the monetary and the effort conditions (black and white, respectively), using only data from the first round. No significant difference in trust was found comparing the monetary and effort conditions (*p-val*= 0. 346; Wilcoxon rank-sum test, n=15). B) Average amount reciprocated by the trustee in the monetary and the effort conditions (black and white, respectively), using only data from the first round. No significant difference was found comparing the monetary and effort conditions (p= 0.68; Wilcoxon rank-sum test, n=15). Error bars represent s.e.m. .

**No gender differences**

To test if there were any gender differences in our results, we compared male and female behavior in the different conditions. For both the monetary and the effort conditions, we found no significant difference between males and females in both the amount trusted ($p\text{-}val_{\$}$=0.45 and $p\text{-}val_{w}$=1; Wilcoxon rank-sum test; $n_{♀}$=17 and $n_{♂}$=13) and the amount reciprocated ($p\text{-}val_{\$}$=1 and $p\text{-}val_{w}$=0.54 for absolute reciprocity; $p\text{-}val_{\$}$=0.37 and $p\text{-}val_{w}$=0.58 for relative reciprocity ; Wilcoxon rank-sum test; $n_{♀}$=18 and $n_{♂}$=12). Also, analyzing separately males and females, we found no differences between behavior in the monetary and the effort condition, in both the amount trusted ($p\text{-}val_{♀}$=0.17 and $p\text{-}val_{♂}$=0.45; Wilcoxon signed-rank test) and the amount reciprocated ($p\text{-}val_{♀}$=0.59 and $p\text{-}val_{♂}$=0.46 for absolute reciprocity; $p\text{-}val_{♀}$=0.41 and $p\text{-}val_{♂}$=0.38 for relative reciprocity; Wilcoxon signed-rank test). This lack of gender differences suggests that the similarity between behavior in the monetary and the effort conditions does not seem to depend on the person's gender.

**Strategies used**

To better understand the strategies used, we graphically analyzed subjects' decisions (Figure 4.6). We observed more variability in the strategies employed in the effort condition. For example, in the effort condition subjects sometimes returned everything or nothing, neither of which happened in the monetary condition. In fact, in the monetary condition the vast majority of the trustees (93%) reciprocated between 1/3 (return exactly what was trusted) and 2/3 (split total earnings), while in the effort condition this percentage, although still high, decays to 70%. All trustors sent something to the trustees, almost all the trustees returned something, and the vast majority of the trustees (93% in the monetary, 87% in the effort condition) returned the same or more than the trustor sent. In several cases, (2 in the effort, 8 in the monetary condition) the trustor sent the entire show-up fee to the trustee, and the trustee returned 2/3 of it back – this point may be considered fair and efficient, since it maximizes the total money/*EB* to be shared and it divides it equitably. Most subjects appear to have followed simple decision-rules in both conditions, and the strategies employed in the effort condition seem more variable.

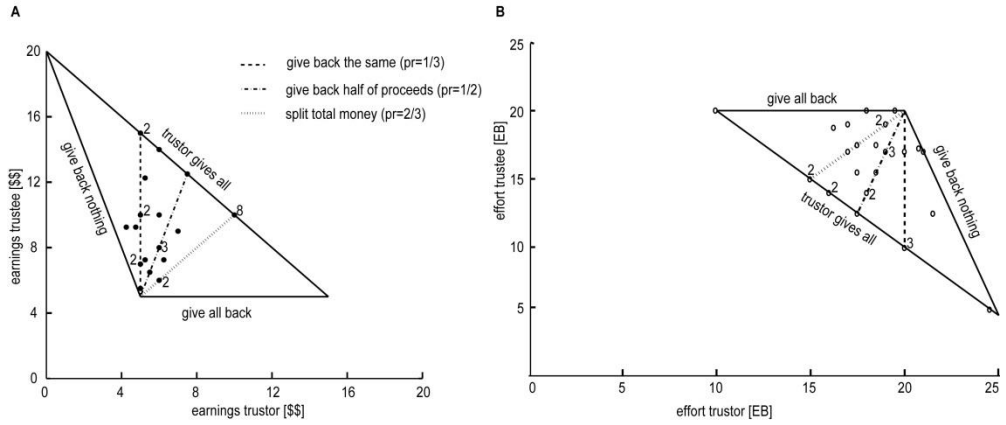


**Figure 4.6 - Distribution of joint earnings or effort. A)** Representation of trustor's and trustee's earnings in the monetary condition(figure similar to the one presented in (Berg et al., 1995)). Each subject pair/output of a round is represented by a filled dot. When dots overlap a small number is shown near the dot, representing the amount of overlapping dots. The outer triangle shows the set of possible earning pairs. The lines represent different possible relative reciprocity values (how much of the tripled amount sent by the trustor the trustee sent back): (1) if the relative return ($p_r$) is equal to zero ("give back nothing" line), it means that the trustee keeps all the money; (2) when $p_r$ =1/3, the trustee sends back the exact amount trusted(dashed line), so the trustor neither wins nor loses. Points that fall to the left of this line indicate that the trustor lost something by trusting, while points at the right of the line indicate the opposite; (3) if $p_r$=1/2, the trustee decides to split the tripled investment in half (dashdot line); (4) when $p_r$ =2/3$, the trustee splits in half the total earnings, inclusive of show-up fees (dotted line); (5) finally , when $p_r$ =1$ the trustee returns the total of the tripled investment, which is the maximum he can return("give all back" line). Dots more near the line confluence vertex of the triangle indicate that the trustor showed lower trust, while dots more near the "trustor gives all" line represent high trust by the trustor. A total number of n=30 subject pairs is represented. **B)** Analogous figure to the one represented in *(A)*, but in which what is represented is the total effort (number of energy blocks that each player had to perform throughout the entire experiment). The triangle here is inverted since increasing trusts decreases the total effort necessary to finish the task (arriving at 20 EB). A total number of n=30 subject pairs is represented.

**Higher behavioral variability in the effort condition**

To further analyze if the behavior is indeed more variable in the effort condition, we computed the dispersion patterns in both conditions (Figure 4.7). Relative reciprocity, although its average does not differ across conditions (recall Figure 4.3), has more variance in the effort condition. Testing for a difference in the variances gives a significant p-value (p=0.002, paired-variance test, n=30). A similar trend can be observed when looking only at the first round (Fig 7 B), although the result is not significant (p=0.19, Bartlett's variance test, n=15). Thus, when subjects traded effort they showed a significantly higher variability of relative reciprocity values.

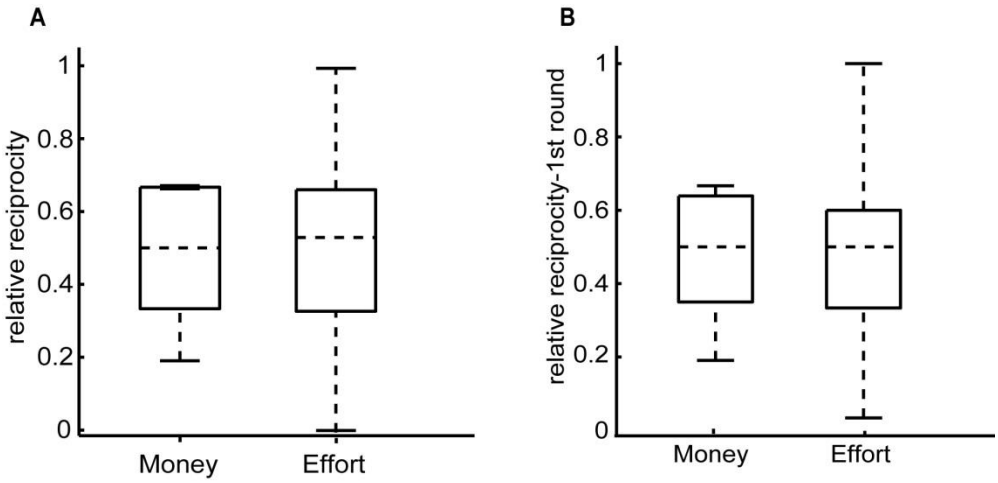


**Figure 4.7 - Box plots of the relative reciprocity obtained for the monetary and effort conditions.** In **A)** all data was used (n=30), while in **B)** just the data concerning the first experiment made by each subject was used (n=15). The middle horizontal dashed line represents the median (50$^{th}$ percentile), the lower horizontal solid line represents the minimum; the bottom and top of the box are the percentile 25$^{th}$ and 75$^{th}$, and the upper horizontal solid line represents the maximum of the data. Note that when no upper black line can be seen it is because the 75$^{th}$ percentile and the maximum are the same.

## 4.5 Discussion

Here we analyzed if trust and reciprocity differ between monetary and effort conditions. Our original hypothesis, based on everyday observations, was that subjects in the effort condition would trust and reciprocate more. However, we found that across the monetary and the physical condition subjects showed the same level of trust and reciprocity. Moreover, subjects' behavior across conditions was highly correlated. The only significant difference was in the variability of behavior, with subjects showing a wider range of relative reciprocity values in the effort condition.

Subjects in both conditions trusted and were trustworthy, sending on average more than half of their monetary earnings or accumulated effort to the other player. The levels of trust and reciprocity observed in the monetary condition are similar to those of analogous monetary experiments reported in the literature (Berg et al., 1995, Fehr et al., 2003). Interestingly, these trusting and reciprocity values were obtained not only when subjects traded money but also when they traded physical effort, suggesting that people's tendency to trust and reciprocate also applies to physical effort decisions.

Previous studies in movement effort decision-making have found some similarities between the decisions made in effort contexts and the ones generally made in monetary settings. For example, in a study in which subjects had to move in order to receive rewards or avoid punishment (Dam, 2009), subjects displayed behaviors such as loss aversion and diminishing returns, phenomena typically described in monetary settings (Maloney et al., 2007). Here, we show that this similarity between movement and economic decisions also extends to social decisions.

Recent research has examined trust and reciprocity in an experiment in which physical effort was involved (Fahr and Irlenbusch, 2000). In a one-shot trust game, subjects had to expend effort, specifically crack walnuts, in order to receive money, which they could then use in a trust game. Both

trustors and trustees tended to give more money to subjects that had worked – it is as if the work had resulted in “property rights”. In this experiment only money was traded and the authors focused on how effort affected the interactions in a trust game, while in our study we asked how the nature of traded units affects behavior and therefore we could ask if trust and reciprocity are shared across these domains.

Our results show, even on the individual level, that trust and reciprocity are very similar between the monetary and the effort conditions. This can indicate that trust and reciprocity may be character traits. This idea is supported by previous studies, which also reported high correlations between cooperative behaviors in a trust game paradigm (meaning trust and trustworthiness) and specific personality traits (Glaeser et al., 1999, Burks et al., 2003, Fehr et al., 2003, Evans and Revelle, 2008). There has been also some indication that part of this cooperative behavior in trust games can be heritable, with monozygotic twins behaving in a more similar way when compared to dizygotic twins (Cesarini et al., 2008). Our results contribute to the view that trust and reciprocity are true character traits.

Why would trust and reciprocity be shared across monetary and effort conditions? Trust and reciprocity have governed social interactions over evolutionary timescales, and it was thus suggested that they could tap into ancient neural systems involved in social cooperation or even directly into reward pathways (McCabe et al., 2001, Rilling et al., 2002, King-Casas et al., 2005, Miller, 2005, Knoch et al., 2006). Since cooperation for joint effort is older than monetary cooperation, we should expect an increased reliance on these primitive pathways. We thus suggest that the behavioral correlation observed between monetary and physical effort conditions may be a result of shared neural substrates. In future studies, this hypothesis could be tested experimentally using neuroimaging approaches.

Another interesting future study to perform would be to check the specific effects of property rights on behavior in the money and effort conditions. In our experiment, while in the effort condition people worked to arrive at 5 energy blocks that they then traded, in the monetary condition people received the $5 as a show up fee without having to work for it. This created a potential difference in the property rights people could feel about the trading units. Surprisingly, this potential difference between conditions did not result in a measurable difference in behavior. It would be interesting to test the effects on behavior when property rights are elicited for the monetary version, for example by having subjects solve a given number of puzzles in order to get the initial $5. We hypothesize that this could create an even higher similarity between behavior in the monetary and the effort conditions.

The only difference we found across the conditions was that relative reciprocity is more variable in the effort condition. This difference can be due to a higher variance of fitness levels across the subject pool compared to variance of average earnings, although this is unlikely the only cause as this higher variability is not exhibited in trusting behavior. It is also possible that reciprocity of effort has less stringent social norms. Given that money is typically easily quantified and carries very strong emotional values (Krueger, 1986), there can be specific social norms on what someone should reciprocate monetarily, but these social norms may be less stringent when it comes to effort retribution. These hypotheses could be tested in future experiments by changing the pool of participants as well as the social framing of the experiment.

The fact that the behavior in both effort and monetary conditions was similar and correlated has two major methodological implications. First, it validates the use of the typical monetary trust game as an effective tool to study trust and reciprocity. Second, it opens the possibility of studying trust and reciprocity using physical effort tasks. Effort based tasks may have a

number of advantages. One advantage is that such experiments may be done cheaply over a wide range of investments (from a single squat to hours of hard workout). Furthermore, it can be a good alternative when comparing trust games across different countries, with different monetary units and/or different purchasing power, as the value of each energy block is less likely to be influenced by the country from which a person is from. Finally, it can allow for a better sampling of the population, as it is easier (and probably cheaper) to get a wider coverage of the population's physical effort cost functions than of the monetary cost functions, and can thus offer us a bigger and more representative set of behaviors. The physical effort condition that we introduced here may be seen as a new tool to study trust and reciprocity, complementing the use of monetary trust games.

### 4.6 Acknowledgements

The authors wish to thank Hugo Fernandes, Asier Mariscal and Karl Mendoza for helpful comments on the manuscript and the participants at the Society for Neuroeconomics meeting (2009) for discussion and suggestions.

### 4.7 References

Berg J, Dickhaut J, McCabe K (1995) Trust, Reciprocity, and Social History. Games and Economic Behavior 10:122-142.

Burks SV, Carpenter JP, Verhoogen E (2003) Playing both roles in the trust game. Journal of Economic Behavior & Organization 51:195-216.

Camerer CF (2003) Behavioural studies of strategic thinking in games. Trends in Cognitive Sciences 7:225-231.

Camerer CF, Fehr E (2002) Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists. SSRN eLibrary.

Cesarini D, Dawes CT, Fowler JH, Johannesson M, Lichtenstein P, Wallace B (2008) Heritability of cooperative behavior in the trust game. P Natl Acad Sci USA 105:3721-3726.

Chami R, Fullenkamp C (2001) Trust and Efficiency. SSRN eLibrary.

Dam G (2009) Learning to move for rewards. vol. PhD dissertation Chicago: Northwestern Universit.

Dirks KT (1999) The effects of interpersonal trust on work group performance. J Appl Psychol 84:445-455.

Dirks KT, Ferrin DL (2001) The role of trust in organizational settings. Organization Science 12:450-467.

Evans AM, Revelle W (2008) Survey and behavioral measurements of interpersonal trust. Journal of Research in Personality 42:1585-1593.

Fahr R, Irlenbusch B (2000) Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment. Economics Letters 66:275-282.

Fehr E, Gachter S (2000) Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives 14:159-181.

Fehr E, Gachter S, Kirchsteiger G (1997) Reciprocity as a contract enforcement device: Experimental evidence. Econometrica 65:833-860.

Fehr E, List JA (2004) The Hidden Costs and Returns of Incentives—Trust and Trustworthiness Among CEOs. Journal of the European Economic Association 2:743-771.

Fehr E, Urs F, Bernhard von R, Juergen S, Gert GW (2003) A Nation-Wide Laboratory. Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Survey. CESifo Group Munich.

Frey BS (1993) Does Monitoring Increase Work Effort - the Rivalry with Trust and Loyalty. Economic Inquiry 31:663-670.

Gachter S, Falk A (2002) Reputation and reciprocity: Consequences for the labour relation. Scandinavian Journal of Economics 104:1-26.

Gambetta D (2000) Can We Trust Trust? In: Trust: Making and Breaking Cooperative Relations(Gambetta, D., ed): Department of Sociology, University of Oxford.

Glaeser EL, Laibson D, Scheinkman JA, Soutter CL (1999) What is Social Capital? The Determinants of Trust and Trustworthiness. Harvard - Institute of Economic Research.

Guerra GA (2002) Crowding out Trust: The Adverse Effect of Verification: An Experiment. SSRN eLibrary.

Guseva A, Rona-Tas A (2001) Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared. American Sociological Review 66:623-646.

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: Reputation and trust in a two-person economic exchange. Science 308:78-83.

Klimoski RJ, Karol BL (1976) Impact of Trust on Creative Problem-Solving Groups. J Appl Psychol 61:630-633.

Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. Quarterly Journal of Economics 112:1251-1288.

Knoch D, Gianotti LRR, Pascual-Leone A, Treyer V, Regard M, Hohmann M, Brugger P (2006) Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. J Neurosci 26:6469-6472.

Krueger DW (1986) Money, success, and success phobia. In: The last taboo: Money as a symbol and reality in psychotherapy and psychoanalysis(Krueger, D. W., ed), pp 3-16 New York: Brunner/Mazel.

Krueger F, McCabe K, Moll J, Kriegeskorte N, Zahn R, Strenziok M, Heinecke A, Grafman J (2007) Neural correlates of trust. Proc Natl Acad Sci U S A 104:20084-20089.

Lauterbach D, Truong H, Shah T, Adamic L (2009) Surfing a Web of Trust: Reputation and Reciprocity on CouchSurfing.com. In: Computational Science and Engineering, 2009 CSE '09 International Conference on, vol. 4, pp 346-353.

Levitt SD, List JA (2007) What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?

Maloney LT, Trommershäuser J, Landy MS (2007) Questions Without Words: A Comparison Between Decision Making Under Risk and Movement Planning Under Risk. In: Integrated Models of Cognitive

Systems(Gray, W. D., ed), pp 297-313 New York: Oxford University Press.

McCabe K, Houser D, Ryan L, Smith V, Trouard T (2001) A functional imaging study of cooperation in two-person reciprocal exchange. Proc Natl Acad Sci U S A 98:11832-11835.

Meyerson D, Weick KE, Kramer RM (1996) Swift trust and temporary groups. In: Trust in organizations: Frontiers of theory and research(Kramer, R. M. and Tyler, T. R., eds), pp pp. 166-195 Thousand Oaks, CA, US: Sage Publications.

Miller G (2005) Neuroscience - Economic game shows how the brain builds trust. Science 308:36-36.

Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437:1291-1298.

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002) A neural basis for social cooperation. Neuron 35:395-405.

van Dijk F, Sonnemans J, van Winden F (2001) Incentive systems in a real effort experiment. European Economic Review 45:187-214.

Williams KD, Karau SJ (1991) Social Loafing and Social Compensation - the Effects of Expectations of Coworker Performance. J Pers Soc Psychol 61:570-581.

# 5. FINAL DISCUSSION

"*It is the uncertainty that charms one. A mist makes things wonderful.*"

*Oscar Wilde*

# 5. FINAL DISCUSSION

This thesis set itself to a special endeavor: reducing the uncertainty in uncertainty. Not that there is anything wrong with it: uncertainty is everywhere and, indeed, a little bit of it makes things interesting. Nevertheless, we spend a lot of time thinking (and sometimes worrying!) about decisions that have uncertainty associated, and so it is kind of comforting to know that our brain has mechanisms to deal with it in quite an acceptable way.

## 5.1 This thesis: brief overview and conclusions

### 5.1.1 Neural correlates of Bayesian statistics

We started by asking which brain areas represent uncertainty in both prior beliefs and current sensory information (likelihood). We found, in chapter 2, that both prior and likelihood uncertainty showed specific neural correlates. Moreover, we found that they have very distinct, non-overlapping representations: Higher prior uncertainty was associated with greater activations in specialized brain areas including the putamen, amygdala, insula and OFC. Likelihood uncertainty, in contrast, was associated with increased activations in higher level visual areas inside the traditional visio-motor pathway. Our findings suggest that, besides behaviorally, prior and likelihood uncertainty can also be distinguished neuronally.

We also found, in chapter 2, that activity from the putamen was not only correlated with increased prior uncertainty but also with subjects' sensitivity to differences in uncertainty and average reliance on current vs.

prior information. This indicates that the putamen may play a special role in decision-making under uncertainty. However, which specific role was unclear: it could be involved in directly encoding prior uncertainty, in signaling the weight given to current vs. prior information, or in attending to the current information. Moreover, as any fMRI data, our imaging results are correlational. In order to try to infer the specific causal role of putamen in decision-making under uncertainty we then performed the experiment reported in chapter 3.

We discovered that Parkinson's disease (PD) patients, who due to their condition have deficits in dopaminergic transmission from the putamen, were impaired in decision-making under uncertainty. Some aspects remained conserved: PD patients were still able to learn the different prior uncertainties as well as controls, and did not show significant differences in the average weight given to current vs. prior information. However, PD patients were impaired at reacting to changes in current sensory uncertainty, and this impairment was especially noticeable when the patients were *off* dopaminergic medication. Our findings suggest that dopaminergic activity from the putamen may be essential for processing uncertainty in current sensory information.

So what may be the specific role of putamen? From both our fMRI and patient results, we suggest that its function may be to promote the person to seek and attend to new incoming sensory information. In this way the putamen would be more active when the prior is more uncertain, signaling that more information has to be gathered. In contrast, less capacity for putamen activation (as in the case of PD patients) would result in less attention to the current stimulus, and hence less responsiveness to its characteristics (such as its uncertainty). This is also in agreement with previous literature. A need for increased gathering and attention to new stimuli should be more prevalent when the information is ambiguous and/or is still being learned, all cases in which putamen activity was found to be higher (Grafton et al., 1995, Hsu et al., 2005, Lehericy et al., 2005, Poldrack

et al., 2005, Doyon et al., 2009, Orban et al., 2010). Furthermore, anatomically the putamen (together with the rest of the basal ganglia) serves as a relay area, receiving inputs from sensory areas and projecting to motor effector systems in the brainstem and cortex (Albin et al., 1989, Whalen, 1998, Rolls, 2000, Singer et al., 2009), and hence it is in an ideal position to facilitate behavioral change to gather relevant new information (Whalen, 1998). Finally, the putamen is also closely connected with reward areas, and could thus easily combine the need for more information with its potential relevance (Rolls, 2000, Preuschoff et al., 2006, Rangel et al., 2008).

### 5.1.2 Role of dopamine in decision-making under uncertainty

The results obtained in chapter 3 can also give us insights into the role of dopamine in decision-making under uncertainty. Indeed, the main pathological hallmark of Parkinson's disease is a pronounced loss of dopamine-producing neurons in the substantia nigra pars compacta, which leads to a drastic depletion of dopamine in the striatum, to which these neurons project (Lotharius and Brundin, 2002). Hence, the differences we observed between PD patients and age-matched controls are likely related to an abnormal dopaminergic transmission (Riederer and Wuketich, 1976). A potential explanation of why dopamine would be affecting the processing of uncertainty in current sensory information is if it drives the subjects to attend to the current information and, by doing so, the subjects also notice its uncertainty more. In contrast, if the subjects are not attending to the current information, they will rely on it in a similar way regardless of its specific uncertainty. This is in accordance with the interpretation that attention is the inference about the uncertainty (or conversely, the precision) of a sensory signal (Feldman and Friston, 2010). It is also in agreement with previous literature involving dopamine in attention and saliency of a stimulus (Berridge and Robinson, 1998, Kapur, 2003, McClure et al., 2003).

According to this view, if a sensory stimulus is novel, intense or potentially important for an animal's fitness, then it will activate a dopaminergic response, and this will lead the animal to direct attention to it (Redgrave and Gurney, 2006). We propose that this increased attention fomented by dopamine will also enable the animal to better detect the uncertainty in the current sensory stimulus, and hence respond appropriately to it.

Another interpretation by why dopamine would be affecting the processing of current sensory uncertainty (chapter 3) could be that it directly codes for the uncertainty (or its inverse, precision) in the current stimulus (Friston et al., 2012). However, under these considerations we would expect PD patients, when compared to controls, to have a generalized increase/decrease reliance on current information, which we do not find. Furthermore, based on a technique that allows distinguishing between sensory perception and attentional mechanisms, a previous study found that PD patients had a specific loss of attentional mechanisms that was independent of low level sensory processing deficits (Sampaio et al., 2011). Also, if it directly codes for the uncertainty in the current stimulus one would expect the brain areas more related with major dopaminergic pathways to be more active with increased current sensory uncertainty, and that is not what we observed in chapter 2. Together, these results indicate that dopamine may not directly code for uncertainty in the current stimulus but, instead, help in signaling its uncertainty by increasing attention to it.

### 5.1.3 Dealing with social uncertainty as a potential character trait

In chapter 4 we found that subjects playing a trust game trusted and reciprocated this trust, and that they did it regardless of what was being traded, money or physical effort. Moreover, we found that there was a significant positive correlation in the way subjects behaved across the two situations. Trust, especially in an anonymous one shot-interaction (in which no specific information from the other player is known), can be considered a

way of measuring how someone deals with social uncertainty (Gambetta, 2000, Guseva and Rona-Tas, 2001). The fact that, even on the individual level, trust was similar between monetary and physical effort conditions indicates that the way people deal with social uncertainty may be a character trait.

Together with the results of the other chapters, these data suggest that there may be individual bias in how people respond to a given type of uncertainty. For example, when dealing with uncertainty in non-social current and prior information, subjects displayed different general tendencies in how much they decided to rely on current vs. prior information, which was shown by their different average Bayesian slopes (chapter 2 and 3). Moreover, these tendencies had a neural correlate: they correlated with how much, on average, subjects' putamen was differentially active in the task (chapter 2). Furthermore, when dopaminergic activity was compromised (as is the case with PD patients) we observed a diminished behavioral sensitivity to differences in current sensory uncertainty (chapter 3). This is in agreement with previous research: for example, anxious people and animals react differently to uncertainty, and these differences may be related to specific neuromodulators (Nader and Balleine, 2007, Tsetsenis et al., 2007). The results presented in this thesis reinforce the idea that there may be individual traits in how someone deals with uncertainty, be it social or non-social, and that these may have a neuronal explanation.

### 5.2 Limitations of the work and future research

Although we found specific brain correlates for prior and likelihood uncertainty, it is possible that uncertainty may be represented differently for other types of tasks. For example, one question that could be asked is if likelihood uncertainty would still be represented in high-level visual areas if the uncertainty was not explicitly obtained from the characteristics of the stimulus *per se* (in this case, dot spread). To answer this question, we will use a similar version of the task utilized in chapter 2 and 3 but in which likelihood uncertainty is provided implicitly. This will be achieved by having only one likelihood dot, have this likelihood dot display different colors depending on the uncertainty of the likelihood information, and have subjects learn the color-uncertainty correspondence. In this way, the dot still transmits information about the likelihood uncertainty, but this uncertainty is not directly related with the characteristics of the stimulus itself (one color *per se* will not be more or less uncertain than the other). We can then compare trials in which likelihood uncertainty is provided explicitly (high vs. lower spread of dots) with trials in which likelihood uncertainty is provided implicitly (learned correspondence between a color and a given uncertainty level), and see if they activate different brain regions.

The prior used in the task from chapters 2 and 3 required learning and memory. Indeed, although subjects were explicitly told the mean of the prior ("the thrower is aiming the coin towards the middle"), they were not told the variance/uncertainty in it, and had to learn it from experience. It is possible that a very different brain activation pattern would be seen if the prior uncertainty was told/shown explicitly and did not need to be learned through experience (d'Acremont and Bossaerts, 2012). To address this question, we will have subjects perform a similar task but in which the prior is given explicitly and does not need to be learned, for example by having the distribution shown in the background.

It would be interesting to further investigate the role of putamen in decision-making under uncertainty. From the results of chapter 2 alone we

were initially expecting prior uncertainty processing to also be affected, which was not what we observed. It is possible that putamen activity still has a role in prior uncertainty processing, but that is independent from dopamine. A possible future study that could elucidate this role further would be to administer a similar task to other patient populations that also have deficits in transmission from the putamen, but in which these deficits are not directly related with a lack of dopamine, such as patients with Wilson's disease (King et al., 1996).

We considered the effects of dopamine, but it is possible (and even likely) that other neurotransmitters may also be involved in decision-making under uncertainty. For example, acetylcholine has been implicated in signaling expected uncertainty, while norepinephrine would signal unexpected uncertainty (Yu and Dayan, 2005). It is also quite possible that different concentrations of the same neuromodulator will give rise to different effects (Cools and Robbins, 2004). Future research could examine the effect of each neuromodulator and its concentration in decision-making under uncertainty, either directly by administration of the corresponding neuromodulator in different quantities, or indirectly by looking at genetic polymorphisms.

Another exciting area of investigation would involve understanding how the way people deal with uncertainty generalizes. It has been found, in a visio-motor rotation task, that changes to prior uncertainty in one reaching direction affect the perception of prior uncertainty in other directions, i.e. it generalizes, even though no training occurred in those directions (Fernandes et al., 2012). It was also found that prior uncertainty in that visuo-motor rotation task generalizes more broadly than the mean of the prior (Fernandes et al., 2012). It would be interesting to know if the way people react to uncertainty in a visual decision-making task (such as ours) can generalize to how they react to uncertainty in other domains, for example social domains, and what are the neural substrates involved.

### 5.3 Potential implications

The results obtained in this thesis help uncover the neural mechanisms behind human prior belief formation (d'Acremont and Bossaerts, 2012), and give insights on how the human brain is capable of performing decisions close to the optimal prescribed by Bayesian statistics. This can have implications for understanding learning disabilities and abnormal decision-making behavior. Learning, from a Bayesian point of view, can be seen as the weight given to new sensory information over prior beliefs (Courville et al., 2006). When the prior is more certain then less learning from new information should occur, and vice versa (Courville et al., 2006). Furthermore, optimal decision-making also implies an accurate evaluation of the uncertainty in both prior beliefs and current information (Kording and Wolpert, 2004, Kording, 2007). Changes in the brain areas or neuromodulators that are involved in this process may lead to an underestimation or overestimation of uncertainty, and the resulting learning or decision-making deficits can be predicted if one knows the particular function of each of these areas and neuromodulators. This knowledge can then help guide targeted solutions, for example by providing explicit information about the actual uncertainty or by incentivizing attention to it.

Our findings can also help guide future models of how the brain represents uncertainty. Many theories have been proposed on how the brain could represent uncertainty (see subsection 1.5 of this thesis). Our findings suggest different types of coding for different types of uncertainty: prior uncertainty seems to be encoded in specialized brain areas (Hsu et al., 2005, Preuschoff et al., 2006, Schultz et al., 2008, Singer et al., 2009), while likelihood uncertainty appears to be encoded in the same neurons that also transmit the stimulus information itself, in this case visual stimulus (Ma et al., 2006, Fiser et al., 2010). Furthermore, dopamine seems to affect the processing of likelihood uncertainty, and given its neuromodulatory property it may be that it influences the strength of the connections between neurons (Wu and Amari, 2003). Future theories about neural representations of

uncertainty should consider prior and likelihood uncertainty separately, given that both neuronally and behaviorally they are separable.

The proposed role of dopaminergic transmission from the putamen in guiding seeking and attention to new incoming sensory information, including to its uncertainty, can aid in the understanding of some of the symptoms PD patients have. Never medicated PD patients have shown selective deficits in novelty seeking, i.e. they seem to have less of a drive to seek new information (Bodi et al., 2009). Interestingly, these deficits can be remediated with dopamine agonist administration (Bodi et al., 2009). Dopamine agonist administration, however, can have its own side effects: it has been reported that it leads to compulsive shopping, hypersexuality and compulsive gambling (Bostwick et al., 2009, Vilas et al., 2012), which can be considered an excessive drive to seek new sensory information. If these symptoms are indeed related to a diminished or excessive attention to new sensory information, then these deficits can be ameliorated by for example increasing the uncertainty in the prior or in the current information.

The observation that the way people trust and reciprocate in a similar way regardless of money or physical effort being exchanged has two main methodological implications: first, it validates the use of the classic monetary trust game as a potential way to measure general trust and reciprocity. Second, it offers the possibility of studying trust and reciprocity in a game in which only physical effort is exchanged. This has potential advantages: a main one is that it would be very easy to change the value of each trading unit (a small change in a single line code would make one "energy block" much more effort demanding), and hence in a much more inexpensive way allow for the study of how trust and reciprocity are affected when there are high stakes at play. Another potential advantage is that it may allow for a better sampling of the population, given that it is easier to obtain a wider coverage of the population's physical effort cost functions than of the monetary cost functions (e.g. it is probably easier to obtain a very fit subject than a very rich one), which can lead to a broader and more

representative set of behaviors. The physical-effort based trust game developed here can thus provide a potential new tool to inexpensively study trust and reciprocity over a wide range of stake sizes and cost functions.

The finding of individual traits and potential neural correlates on how someone deals with uncertainty can have implications for anxiety disorders and depression. Anxious people and animals overreact to uncertain or ambiguous sensory stimuli (Nader and Balleine, 2007, Tsetsenis et al., 2007) and a biased negative interpretation of uncertain stimuli may be a key factor leading to depression (Beck, 2008, Enkel et al., 2010). Interestingly, these behaviors have been associated with an overreactive amygdala (Tsetsenis et al., 2007, Enkel et al., 2010), which is precisely one of the brain areas we found more active with increased prior uncertainty, i.e. uncertainty in one's beliefs. Anxiety and depression can thus in part result from an incorrect estimation of prior and current sensory uncertainty due to under or overreactivity of the corresponding brain areas.

### 5.4 Conclusion

Taken together, the results from this thesis suggest that we are, in general, quite well equipped with dealing with uncertainty. We can take into account uncertainty in both prior and current sensory information. We can even combine these different sources of uncertainty in a way that is qualitatively close to the optimum prescribed by Bayesian statistics. We seem to do so by devoting specific brain areas to each kind of uncertainty: prior uncertainty was correlated with increased activations in putamen, amygdala, insula and OFC, while current sensory uncertainty was associated with activations in higher-level visual areas (chapter 2). Moreover, some particular brain areas and neuromodulators, namely putamen and dopamine, potentially tell us when we should pay attention to current sensory information and its uncertainty, and when not (chapter 2 and 3). Finally, when faced with social

uncertainty, we seem to trust in a similar way regardless of what is being traded, money or physical effort (chapter 4). And this trust pays off: subjects in our experiment on average trusted their anonymous partners, and received more than if they had not trusted at all.

This thesis was a small step towards the understanding of decision-making under uncertainty and the neural mechanisms involved in it. Much still needs to be researched about this topic: for example, what are the causal roles of the other brain areas we found related with prior uncertainty? How generalizable are these results? Are there other neuromodulators involved? How do these brain areas and neuromodulators interact? The answer to these and other questions is still uncertain. And that is wonderful.

## 5.5 References

Albin RL, Young AB, Penney JB (1989) The Functional-Anatomy of Basal Ganglia Disorders. Trends Neurosci 12:366-375.

Beck AT (2008) The evolution of the cognitive model of depression and its neurobiological correlates. Am J Psychiat 165:969-977.

Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Rev 28:309-369.

Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibo G, Takats A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinsons patients. Brain 132:2385-2395.

Bostwick JM, Hecksel KA, Stevens SR, Bower JH, Ahlskog JE (2009) Frequency of New-Onset Pathologic Compulsive Gambling or Hypersexuality After Drug Treatment of Idiopathic Parkinson Disease. Mayo Clin Proc 84:310-316.

Cools R, Robbins TW (2004) Chemistry of the adaptive mind. Philos T Roy Soc A 362:2871-2888.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences 10:294-300.

d'Acremont M, Bossaerts P (2012) Decision Making: How the Brain Weighs the Evidence. Current Biology 22:R808-R810.

Doyon J, Bellec P, Amsel R, Penhune V, Monchi O, Carrier J, Lehericy S, Benali H (2009) Contributions of the basal ganglia and functionally

related brain structures to motor learning. Behavioural Brain Research 199:61-75.

Enkel T, Gholizadeh D, von Bohlen und Halbach O, Sanchis-Segura C, Hurlemann R, Spanagel R, Gass P, Vollmayr B (2010) Ambiguous-Cue Interpretation is Biased Under Stress- and Depression-Like States in Rats. Neuropsychopharmacol 35:1008-1015.

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4.

Fernandes HL, Stevenson IH, Kording K (2012) Generalization of uncertainty. In: Computational and Systems Neuroscience Salt Lake City.

Fiser J, Berkes P, Orban G, Lengyel M (2010) Statistically optimal perception and learning: from behavior to neural representations. Trends in Cognitive Sciences 14:119-130.

Friston KJ, Shiner T, FitzGerald T, Galea JM, Adams R, Brown H, Dolan RJ, Moran R, Stephan KE, Bestmann S (2012) Dopamine, Affordance and Active Inference. Plos Comput Biol 8.

Gambetta D (2000) Can We Trust Trust? In: Trust: Making and Breaking Cooperative Relations(Gambetta, D., ed): Department of Sociology, University of Oxford.

Grafton ST, Hazeltine E, Ivry R (1995) Functional Mapping of Sequence Learning in Normal Humans. J Cognitive Neurosci 7:497-510.

Guseva A, Rona-Tas A (2001) Uncertainty, Risk, and Trust: Russian and American Credit Card Markets Compared. American Sociological Review 66:623-646.

Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

Kapur S (2003) Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiat 160:13-23.

King AD, Walshe JM, Kendall BE, Chinn RJ, Paley MN, Wilkinson ID, Halligan S, Hall-Craggs MA (1996) Cranial MR imaging in Wilson's disease. AJR Am J Roentgenol 167:1579-1584.

Kording K (2007) Decision theory: what "should" the nervous system do? Science 318:606-610.

Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. Nature 427:244-247.

Lehericy S, Benali H, Van de Moortele PF, Pelegrini-Issac M, Waechter T, Ugurbil K, Doyon J (2005) Distinct basal ganglia territories are engaged in early and advanced motor sequence learning. P Natl Acad Sci USA 102:12566-12571.

Lotharius J, Brundin P (2002) Pathogenesis of Parkinson's disease: Dopamine, vesicles and alpha-synuclein. Nat Rev Neurosci 3:932-942.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nature Neuroscience 9:1432-1438.

McClure SM, Daw ND, Montague PR (2003) A computational substrate for incentive salience. Trends Neurosci 26:423-428.

Nader K, Balleine B (2007) Ambiguity and anxiety: when a glass half full is empty. Nature Neuroscience 10:807-808.

Orban P, Peigneux P, Lungu O, Albouy G, Breton E, Laberenne F, Benali H, Maquet P, Doyon J (2010) The multifaceted nature of the relationship between performance and brain activity in motor sequence learning. Neuroimage 49:694-702.

Poldrack RA, Sabb FW, Foerde K, Tom SM, Asarnow RF, Bookheimer SY, Knowlton BJ (2005) The neural correlates of motor skill automaticity. J Neurosci 25:5356-5364.

Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. Neuron 51:381-390.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545-556.

Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nat Rev Neurosci 7:967-975.

Riederer P, Wuketich S (1976) Time Course of Nigrostriatal Degeneration in Parkinsons-Disease - Detailed Study of Influential Factors in Human-Brain Amine Analysis. J Neural Transm 38:277-301.

Rolls ET (2000) The orbitofrontal cortex and reward. Cerebral Cortex 10:284-294.

Sampaio J, Bobrowicz-Campos E, André R, Almeida I, Faria P, Januário C, Freire A, Castelo-Branco M (2011) Specific impairment of visual spatial covert attention mechanisms in Parkinson's disease. Neuropsychologia 49:34-42.

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, Bossaerts P (2008) Explicit neural signals reflecting reward uncertainty. Philosophical Transactions of the Royal Society B-Biological Sciences 363:3801-3811.

Singer T, Critchley HD, Preuschoff K (2009) A common role of insula in feelings, empathy and uncertainty. Trends in Cognitive Sciences 13:334-340.

Tsetsenis T, Ma XH, Lo Iacono L, Beck SG, Gross C (2007) Suppression of conditioning to ambiguous cues by pharmacogenetic inhibition of the dentate gyrus. Nature Neuroscience 10:896-902.

Vilas D, Pont-Sunyer C, Tolosa E (2012) Impulse control disorders in Parkinson's disease. Parkinsonism Relat Disord 18 Suppl 1:S80-84.

Whalen PJ (1998) Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. Curr Dir Psychol Sci 7:177-188.

Wu s, Amari S (2003) Neural Implementation of Bayesian Inference in Population Codes. Advances in Neural Information Processing Systems 2003:1-8.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.

# 6. APPENDIX

## 6. APPENDIX

### Supplementary tables

| cluster | | | Voxel | | | | | | Regions |
|---|---|---|---|---|---|---|---|---|---|
| pcor[1] | Ke[2] | punc[3] | pcor[4] | T[5] | Z[6] | x[7] | y[7] | z[7] | Region - L/R[8] |
| 0 | 114 | 0 | 0.001 | 7.27 | 5.82 | -30 | 8 | -26 | temp pole-L |
| | | | 0.95 | 4.25 | 3.86 | -22 | 0 | -22 | Amygdala-L |
| 0.02 | 40 | 0.004 | 0.001 | 7.1 | 5.72 | 24 | -18 | 12 | putamen/ i.c. -R |
| 0.124 | 21 | 0.029 | 0.001 | 7.08 | 5.71 | 58 | 6 | -24 | temp mid -R |
| 0.001 | 86 | 0 | 0.002 | 6.83 | 5.57 | 30 | -2 | -2 | putamen-R |
| | | | 0.056 | 5.74 | 4.91 | 20 | -6 | -8 | g pallidus-R |
| 0.014 | 44 | 0.003 | 0.007 | 6.5 | 5.38 | 30 | 16 | -30 | temp pole-R |
| 0.001 | 72 | 0 | 0.015 | 6.21 | 5.2 | 62 | 28 | -2 | front inf tri-R |
| | | | 0.261 | 5.14 | 4.5 | 66 | 14 | 4 | front inf oper-R |
| 0.003 | 62 | 0.001 | 0.018 | 6.16 | 5.17 | 16 | -42 | 52 | PCC/ pc lob-R |
| | | | 0.097 | 5.54 | 4.77 | 18 | -30 | 40 | PCC-R |
| 0.112 | 22 | 0.026 | 0.024 | 6.05 | 5.1 | -34 | -20 | 10 | insula-L |
| 0.191 | 17 | 0.046 | 0.029 | 5.98 | 5.06 | 66 | 2 | -12 | temp sup-R |
| 0.082 | 25 | 0.019 | 0.042 | 5.84 | 4.97 | 20 | -4 | 14 | between putamen and caudate/ i.c. (genu)-R |
| 0.074 | 26 | 0.017 | 0.045 | 5.82 | 4.96 | -60 | -8 | -14 | STS-L |
| 0.034 | 34 | 0.008 | 0.048 | 5.8 | 4.94 | 22 | 26 | -12 | OFC-R |
| 0.004 | 60 | 0.001 | 0.106 | 5.5 | 4.75 | 22 | 0 | -20 | amygdala-R |
| 0.034 | 34 | 0.008 | 0.189 | 5.28 | 4.59 | -56 | 8 | -28 | temp mid-L |
| | | | 0.273 | 5.12 | 4.49 | -52 | 0 | -32 | temp mid-L |
| 0.031 | 35 | 0.007 | 0.195 | 5.26 | 4.59 | -44 | 10 | -18 | temp pole sup-L |
| 0.411 | 10 | 0.115 | 0.275 | 5.12 | 4.49 | -22 | 0 | 4 | putamen-L |

**Supplementary Table S1. Areas more active with increased prior uncertainty.** Local maxima of clusters, $p < 0.0001$ uncorrected. Clusters with $k < 10$ voxels not shown. Columns of the table represent, by order: [1]Corrected (family-wise) cluster-level *p*-value; [2]Cluster size (voxels); [3]Uncorrected cluster-level *p*-value; [4]Corrected (family-wise) voxel-level *p*-value; [5]T-statistic of voxel; [6]Z-score of voxel; [7](x, y, z) are the MNI coordinate of voxel location (mm); [8]Laterality (L = left, R = right). Regions legend: front inf oper = inferior frontal (pars opercularis); front inf tri = inferior frontal (pars triangularis); g pallidus = globus pallidus; i.c. = internal capsule; pc lob = paracentral lobule; PCC = posterior cingulate cortex; STS = superior temporal sulcus; temp mid= middle temporal; temp pole = temporal pole; temp sup = temporal superior.

| cluster | | | Voxel | | | | | | Regions |
|---|---|---|---|---|---|---|---|---|---|
| **pcor** | **Ke** | **punc** | **pcor** | **T** | **Z** | **x** | **y** | **z** | **Region – L/R** |
| 0.009 | 49 | 0.002 | 0 | 10.13 | 7.17 | -16 | 30 | 14 | near ACC-L |
| 0 | 217 | 0 | 0 | 7.89 | 6.14 | 22 | 26 | 28 | front sup-R |
| | | | 0.004 | 6.64 | 5.46 | 14 | 24 | 28 | ACC-R |
| | | | 0.032 | 5.95 | 5.04 | 24 | 24 | 36 | front mid-R |
| 0.02 | 40 | 0.004 | 0 | 7.37 | 5.87 | 6 | 22 | 10 | near caudate-R |
| | | | 0.591 | 4.73 | 4.21 | 10 | 14 | 6 | caudate-R |
| 0.018 | 41 | 0.004 | 0.001 | 7.1 | 5.72 | 36 | 6 | 26 | front inf ope-R |
| 0.091 | 24 | 0.021 | 0.003 | 6.74 | 5.52 | 24 | 10 | 58 | front sup-R |
| 0.02 | 40 | 0.004 | 0.003 | 6.73 | 5.51 | -50 | -54 | -2 | temp mid-L |
| 0.002 | 66 | 0.001 | 0.008 | 6.42 | 5.33 | 36 | 18 | -6 | insula-R |
| | | | 0.013 | 6.27 | 5.24 | 30 | 24 | 2 | insula-R |
| 0.002 | 68 | 0 | 0.012 | 6.29 | 5.25 | -28 | -78 | -18 | fusiform-L |
| | | | 0.025 | 6.03 | 5.09 | -18 | -74 | -20 | cerebellum |
| 0 | 89 | 0 | 0.016 | 6.2 | 5.2 | 28 | -92 | 6 | occip mid-R |
| | | | 0.052 | 5.77 | 4.93 | 36 | -82 | 2 | occip mid-R |
| | | | 0.956 | 4.23 | 3.84 | 26 | -86 | 0 | occip mid-R |
| 0.411 | 10 | 0.115 | 0.017 | 6.16 | 5.17 | -20 | -18 | 6 | thalamus-L |
| 0.055 | 29 | 0.012 | 0.018 | 6.16 | 5.17 | 28 | -68 | -24 | cerebellum |
| 0.028 | 36 | 0.006 | 0.021 | 6.11 | 5.14 | 40 | -74 | 40 | angular-R |
| 0.005 | 56 | 0.001 | 0.026 | 6.02 | 5.08 | 42 | -44 | -18 | fusiform-R |
| | | | 0.168 | 5.32 | 4.63 | 52 | -42 | -10 | temp inf-R |
| | | | 0.248 | 5.16 | 4.52 | 52 | -48 | -18 | temp inf-R |
| 0.022 | 39 | 0.005 | 0.047 | 5.81 | 4.95 | -42 | -74 | 18 | occip mid-L |
| 0 | 102 | 0 | 0.049 | 5.79 | 4.94 | 50 | -60 | 2 | temp mid-L |
| | | | 0.155 | 5.36 | 4.65 | 38 | -66 | 8 | occip mid-R |
| 0.034 | 34 | 0.008 | 0.054 | 5.75 | 4.91 | 8 | 22 | 54 | SMA-R |
| | | | 0.526 | 4.8 | 4.26 | 8 | 30 | 48 | front sup med-R |
| 0.05 | 30 | 0.011 | 0.055 | 5.75 | 4.91 | 26 | 64 | 4 | front sup-R |
| 0.038 | 33 | 0.008 | 0.089 | 5.57 | 4.79 | -28 | 34 | 10 | front inf tri-L |

**Supplementary Table S2. Areas more active with decreased prior uncertainty.** Local maxima of clusters, $p < 0.0001$ uncorrected. Clusters with $k < 10$ voxels not shown. Columns of the table are the same as in Supplementary Table 1. Regions legend: ACC = anterior cingulate cortex; front inf oper = inferior frontal (pars opercularis); front inf tri = inferior frontal (pars triangularis); front mid = middle frontal; front sup = superior frontal; front sup med = medial superior frontal; occip mid = middle occipital; SMA = supplementary motor area; temp inf = inferior temporal.

| Contrast | Cluster | | | Voxel | | | | | | Regions |
|---|---|---|---|---|---|---|---|---|---|---|
| | pcor | Ke | punc | pcor | T | Z | x | y | z | Region–L/R |
| **L>l** | 0.002 | 69 | 0 | 0.05 | 5.79 | 4.93 | -14 | -92 | 22 | occip sup-L |
| **L>l** | 0.265 | 14 | 0.067 | 0.851 | 4.44 | 3.99 | 20 | -90 | 18 | occip sup-R |
| **l > L** | 0.295 | 13 | 0.076 | 0.812 | 4.49 | 4.03 | 34 | -88 | 2 | occip mid-R |
| **pL** | 0 | 1770 | 0 | 0.001 | 12.31 | 5.8 | 6 | -90 | 2 | calcarine-R |
| **pL** | | | | 0.001 | 11.97 | 5.74 | 10 | -92 | 14 | cuneus-R |
| **pL** | | | | 0.008 | 10.07 | 5.36 | -20 | -90 | 28 | occip sup-L |
| **pL** | | | | 0.02 | 9.34 | 5.18 | 24 | -88 | 20 | occip sup-R |
| **pL** | 0.234 | 10 | 0.028 | 0.374 | 7.26 | 4.6 | 8 | -76 | 54 | precuneus |
| **pL** | 0.002 | 38 | 0 | 0.809 | 6.38 | 4.3 | 2 | -54 | 52 | precuneus -R |
| **pL** | 0.126 | 13 | 0.014 | 0.928 | 6 | 4.16 | -10 | -70 | 58 | precuneus -L |
| **pBS** | 0.034 | 22 | 0.004 | 0.085 | 8.26 | 4.9 | 12 | 48 | 32 | front sup med-R |

**Supplementary Table S3. Areas more active with increased likelihood uncertainty (L>l contrast), decreased likelihood uncertainty (l> L contrast), a parametric increase in likelihood uncertainty (pL contrast) and a parametric increase in the instantaneous Bayesian slope (pBS contrast).** Local maxima of clusters, $p < 0.0001$ uncorrected. Clusters with $k < 10$ voxels not shown. Columns of the table are the same as in Supplementary Table 1. Region legend: occip sup = superior occipital; occip mid = middle occipital; front sup med = medial superior frontal.